Assessment of the Expert Locomotive Engineer's Mental Model through Expert-Novice Interactions

by

Rachel Price

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

© Massachusetts Institute of Technology 2020. All rights reserved.

Author . . . .
Department of Aeronautics and Astronautics
May 19, 2020

Certified by . . . .
Charles M. Oman
Senior Lecturer and Senior Research Engineer
Thesis Supervisor

Accepted by . . . .
Sertac Karaman
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Assessment of the Expert Locomotive Engineer's Mental Model through Expert-Novice Interactions

by

Rachel Price

Submitted to the Department of Aeronautics and Astronautics
on May 19, 2020, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

Abstract

Today, many long-haul freight locomotives around the world are equipped with autothrottle systems that follow pre-computed and fuel-efficient speed plans. However, these systems cannot adapt to changes in operational constraints or engineers' train handling preferences, which results in engineers taking back manual control. To address issues created by this traded approach scheme, a new operational mode is envisioned that allows operators to shape automation behavior. Although high level goals have been enumerated by previous task analyses, there has been little research on how engineers actually drive routes, identify situations, and make train handling decisions. To fill this gap , five subject pairs drove a U.S. DOT/FRA freight locomotive research simulator along a 65 mile route, responding to signals, speed restrictions and dispatcher orders. Each subject pair consisted of one expert and one novice subject. One subject was seated at the controls and the other subject was seated in the conductor's position. The subject at the controls had limited access to information and relied on verbal communication with the other subject to safely manipulate the train controls. Subjects drove the route twice, once at each position. The research team developed a coding scheme based on cognitive linguistics research and prior work on freight driving strategies to categorize each interaction from the study. Analysis of this data suggested that experienced engineers know what decisions and actions should be taken when various situations are encountered along a route, but their train handling (e.g. braking) tactics vary. Next-generation autothrottle systems should leverage the engineer's ability to assess operational context and initiate actions. Additionally, these systems should allow the operator to make speed plan modifications at both the tactical and strategic level to accommodate the observed variation between engineers' control strategies.

Thesis Supervisor: Charles M. Oman
Title: Senior Lecturer and Senior Research Engineer

Acknowledgments

This thesis came out of a larger work that is funded by the Federal Railroad Administration and is currently being completed by General Electric and Massachusetts Institute of Technology. The experiment described herein took place at the Cab Technology Integration Lab (CTIL) at the Volpe National Transportation Systems Center.

I would like to thank the members of the GE/MIT project team for their efforts in seeing this experiment through. First, to Dr. Chuck Oman, my advisor, for all of his help in the design and execution of the experiment as well as the analysis and presentation of the results in this thesis. Thanks also to Dr. Andy Liu for his guidance and support throughout this project, be it editing the thesis or coordinating with the Volpe Center to make sure the experiments would run smoothly. I am also very grateful for the contributions of Zach Johnson and Laura Rosado, the undergraduate researchers who performed much of the data labelling and assisted with data processing. On the GE side, thanks to Dr. James Brooks, Brad Miller, Neeraja Subrahmaniyan, Kirk Matthews, and Sarah Felix for their assistance and expertise in various subject areas, including the rail industry, cognitive linguistics, and control theory. From the Volpe Center, I would like to acknowledge Matt Isaacs, who was crucial to the successful completion of the experiments.

Finally, I would like to thank George Newman and Alex Goodwin, who helped refine the pilot version of the study, as well as the ten subjects who volunteered for and completed the final study.

This project was supported by DOT/FRA Contract 693JJ6-18-C-000031.

Contents

List of Figures

List of Tables

Acronyms

- **COUHES:** Committee on the Use of Humans as Experimental Subjects
- **CTA:** Cognitive Task Analysis
- **CTIL:** Cab Technology Integration Laboratory
- **EAC:** Engineer-at-the-Controls
- **FRA:** Federal Railroad Administration
- **HTA:** Hierarchical Task Analysis
- **KDP:** Key Decision Point
- **NAC:** Novice-at-the-Controls
- **TO:** General Electric Trip Optimizer
- **TSR:** Temporary Speed Restriction

Chapter 1

Introduction

This thesis describes the methods and execution of a study designed to identify the tactics and strategy used by freight rail engineers when driving manually. The goal was to understand the cues, actions, and context that help form the mental model of an expert engineer. This study was motivated by the need for advanced automation design in locomotive cabs that are intuitive to understand and control. In many ways, the rail industry has been slow to adopt modern automation technologies, especially compared to other forms of transit, such as automobiles or airplanes. While some improvements to in-cab technology have been made in recent years, there are still significant gaps that must be overcome. The study described in this work is part of a larger effort to address these gaps. As part of the design process for in-cab automation system enhancements, it was recognized that the expert engineer's mental model of how to drive the train was not sufficiently well understood. This motivated the present study.

It was not until the mid-to-late 2000s that a commercially successful automatic power management system was introduced for freight locomotives. Touted primarily as a way for railroads to further increase fuel efficiency, multiple automated speed control systems are in use across the U.S today, including the Wabtec Leader platform and the GE Trip Optimizer (TO) platform (Eldredge and Houpt 2011). These automated control systems form one of the most substantial changes to freight locomotive cab operations since the transition from steam to diesel locomotives. Recent accidents have also pushed Congress to mandate Positive Train Control (PTC) on major rail routes, requiring installation of equipment on locomotives and along routes, as well as training rail crew in PTC-enabled operations (American Association of Railroads 2020). Unlike Trip Optimizer, PTC is not designed as a nominal operating mode, but is a safety backup that automatically stops the train in certain hazardous conditions. In addition to improving safety across the national rail system, PTC infrastructure will also enable new cab technologies. For example, the ability to communicate an upcoming signal status to the engineer or in-cab automation means that both the human and the automation can adapt to situations as they evolve, allowing for more flexible control modes. However, with the widespread use of both automated throttle systems and PTC-like safety nets, it is possible that crews will become less skilled as a result of decreased manual operating experience and more complacent due to over-reliance on the safety systems (Roth et al. 2013). These potential changes have spurred interest in further development of new operating modes and automated driver aids that increase both safety and efficiency without diminishing the engineer's skill and role.

This thesis describes early stages of research for the development of a shared control model for freight rail, sometimes referred to as an "enhanced" or "robust" manual mode (Brooks et al. 2016). The hope for this mode is that the engineer is able to continuously adjust the goals of the automated system to achieve safe and efficient management of the train's movement. This contrasts with the current interaction model, where the human operator and automation trade total control back and forth between fully manual modes and fully automated modes, so the operator has no influence on train behavior when not in manual control. For a shared control mode to be effective, the engineer would need a functional mental model of the automation - that is, the ability to understand and predict the behavior of the automation (for example, speed of the train) given his or her inputs to the automation. This could be more readily achieved if the enhanced automation was designed to behave in a manner that reflected the intentions and goals of the human engineer. In other words, building expert driving strategy and experts' mental models into the automation facilitates a shared understanding of the world between operator and automation.

To better understand expert driving strategy, an experiment was conducted in the Cab Technology Integration Laboratory rail cab simulator at the Department of Transportation's Volpe Center in Cambridge, MA. This thesis explains the rationale and methods behind the experiment, the experimental results, and the implications of those results for the robust manual mode design, as well as identifying issues requiring further study.

1.1 Manual Train Control

Modern freight rail locomotives primarily have three means of control: the throttle, the dynamic brake, and the air brake (sometimes referred to as the train brake). The throttle has eight discrete power settings, often called notches, where notch one is the lowest power setting and notch eight is the highest. The engineer can directly influence tractive power through notch, however, the resulting overall train speed is a nonlinear function of many factors, including notch, track incline (grade), track condition, and consist (i.e., the composition and cargo of the entire train) - there is no single notch setting that guarantees a certain speed or speed range. For conventional manual control, this means that the engineer must closely monitor and constantly adjust throttle settings to achieve a desired speed profile. The dynamic brake converts the kinetic energy of the locomotive into an electric current that is then dissipated as heat in resistors atop the locomotive body, thereby slowing the locomotive down (McGonigal 2006). In some cab consoles, the dynamic brake uses the same control interface as the notch. In such cases, there is a single lever that controls both throttle and dynamic brake, with the lever's range divided between a throttle region and a dynamic brake region. The lever is centered in the idle position. Other console designs have separate handles for the dynamic brake and the notch, but with an interlock that prevents both from being used simultaneously. Unlike the dynamic brake, the air brake acts upon the all of the cars in the train.

The principal component of the air brake system is a pipe that runs the length of the train and holds pressurized air that is partially released when the engineer manipulates the air brake control lever in the locomotive. When the pressure in this pipe is lowered, each freight car's brakes are applied by a valve in each wheelset. However, it takes some time for the pressure to equalize along the entire train, so the freight cars at the front of the train will deploy brakes before the rear of the train. The engineer can release further brake pipe air pressure if braking is not sufficient; however, due to the design of the wheelset brake valves, the only way to decrease braking is to repressurize the brake line entirely, which requires entirely releasing all the air brakes. This creates some challenging tactical considerations since the engineer must anticipate how much of time will be needed to recharge the brake pipe pressure in order to make them usable again.

When driving a route, there are many factors for an engineer to consider. First, freight trains vary in length but can be over a mile long, weighing thousands of tons (United States Government Accountability Office 2019). Heavier trains respond more slowly to throttle and brake inputs and in some cases can take more than a minute to change speed. Considerable experience is required to properly anticipate the effects of throttle or brake control inputs. Furthermore, the couplings between train freight cars allow each to move independently over a short distance ("slack action"). Since power can only be applied by the locomotive wheels, increasing the throttle will cause the locomotive to pull away from the cars behind it, and the couplings between successive cars can break if both throttle and brake are not carefully managed. Done correctly, the pulling force from the locomotive propagates down the train ("stretching the train"), setting the entire train into motion. During dynamic braking action, "bunching" of the train can occur, where the locomotive slows down relative to the rest of the train, and the couplings between cars transmit the braking force down the entire length of the train. The air brake can also cause bunching and stretching of the train due to unequal brake application while the brake pipe pressure equalizes. The engineer must also consider the loading distribution of the entire consist, since the behavior of the cars will vary based on their weight. All of these factors significantly complicate freight train handling using throttle, dynamic brake and train brakes, and considerable experience is required to avoid breaking couplers between cars in many common situations.

Engineers must be certified to operate on a particular route, which includes demonstrating that they have memorized the speed limit for each section of track as well as the location of signals along the route. The terrain of the route, specifically the grade, and places where brakes should usually be applied and released must also be memorized. Careful management of the bunching and stretching action is necessary on hills, where the front portion of the train may be accelerating downhill while the back of the train is still working against gravity. This creates high forces on the links of the cars cresting the hill at any given time. Braking strategy is especially difficult given the slow and variable application of the air brake. For reasons discussed above, as the train travels downhill, the engineer must keep the speed below the limiting track speed but at the same time take care not to over-apply the train brake, since it cannot be partially released.

Given these considerations, it clearly requires significant skill and experience to control a freight train safely, and it is even more difficult to perform this task efficiently (e.g., minimizing fuel/energy usage). Any use of either dynamic or train brakes dissipates energy as heat. In theory, there is some minimum braking energy loss associated with driving a given route, defined by grade, speed limits, consist characteristics and track conditions. Additional energy is consumed responding to various signals or other events requiring speed changes. An engineer's driving strategy is defined by the way that he or she chooses to carry out the driving task at both a tactical and strategic level. Tactical decisions affect the execution of a given task, such as how long an engineer remains at a throttle setting before applying more power, bearing in mind that too much power applied too quickly could break the car linkages. In this case, the decision to wait a shorter or longer amount of time is independent of the engineer's larger immediate goal, which is to bring the train up to speed. A strategic decision will be more goal-based: for instance, a strategic decision might be where an engineer chooses to begin slowing down for a red signal. Both types of decisions depend on the engineer's "mental model" for the consist – how it responds to throttle and brake applications, and also the railroad's operating rules that dictate how to respond to wayside signal changes and track speed limits.

Today, automated energy management systems are available to carry out the complex task of managing a freight train's movement and are increasingly used in commercial operation. However, as will be discussed in the following sections, there are several operating conditions where the engineer must still operate the train, thus illuminating the need for more sophisticated automated modes and forming the basis for this project.

Chapter 2

Motivation

2.1 Changes to the Railway Workforce

2.1.1 GE Trip Optimizer and Expert De-Skilling Concerns

In 2009, Houpt et al. described a revolutionary "locomotive control system enhancement", termed Trip Optimizer (TO), that promised to deliver efficiency improvements for rail operations (Houpt et al. 2009). TO consists of two parts: a planning system that pre-computes an optimal speed profile for a given route section that minimizes fuel use given the consist and its distribution, number and type of locomotives, and route characteristics; and a "dynamic control system" that executes the generated speed plan, while incorporating train handling rules to prevent coupler damage. The TO speed profile for the next several miles is shown on a moving map cab display, and allowing the engineer to anticipate the future behavior of the automation. Trip Optimizer saves fuel by reducing unnecessary braking and subsequent accelerations, since braking converts kinetic energy into heat that is dissipated as a waste product. For example, TO's control system might bring the throttle to idle earlier than a human engineer would in order to let air resistance and the track friction slow the train to the appropriate speed. Analysis of TO performance in revenue service showed significant fuel savings (up to 13%) while maintaining train handling comparable to an experienced crew. As of 2020, TO has been widely deployed across the United

States and other countries, with over hundreds of millions of miles driven (Eldredge and Houpt, 2011).

However, this widespread deployment of an automated control mode has raised some concerns that this might lead to a degradation in the manual control skills of the engineers, often referred to as “de-skilling”. Since at least 1983, researchers have noted that higher levels of automation frequently correspond to decay of manual skills in human operators and a lack of appropriate experience. Bainbridge suggested that it is more difficult for humans to effectively supervise and then take over for automation when the human infrequently takes manual control (Bainbrige 1983). In addition to depriving the operator of valuable manual control practice, higher levels of automation can also cause what is dubbed the “out-of-the-loop” problem: an operator supervising an automation system or automated process will be less able to a) identify when system errors occur and b) manually perform tasks after failure occurs. Endsley and Kiris argued that “a loss of situational awareness (SA)” is the primary driver for these phenomena (Endsley and Kiris 1995). The current state of in-cab automation does little to counter either the loss of manual skills or the out-of-the-loop problem, leaving a gap for future automated modes to address.

The problem of de-skilling arises in many different contexts, and a close parallel to the freight rail case can be in found in aircraft cockpit automation. Due to the workload of controlling an aircraft in three dimensions, aircraft cockpits have been automated earlier and more heavily than locomotive cabs. Current cockpit automation is complex and confusing to operate, and its widespread use has been implicated as a factor in multiple recent aviation accidents (Elias 2019). A 2007 study found “considerable knowledge gaps in pilots’ mental models of the automation” when pilots were asked to recover from a simulated disturbance in automation performance. While none of the simulated trials ended in a catastrophic event, many pilots were unable to meet performance objectives for the aircraft (Nikolic and Sarter 2007). As more complicated automation modes are introduced into locomotive cabs, automation designers will need to recognize that the skills of the human operators, over time, may deteriorate and affect their ability to take over from the automation.

2.1.2 An Aging Workforce

In addition to the concerns of de-skilling due to increased automation, the rail industry is expected to suffer large knowledge and experience losses as the current workforce begins to retire. A 2011 Federal Railroad Administration (FRA) report projected that nearly half of the railroad employees in the United States would be eligible for retirement in 2019. The replacement engineers may not develop the same level of expertise as their predecessors if they predominantly operate in the automated modes and may have more difficulty taking control from the automation. In addition to gaps left by retiring employees, the FRA projects more jobs becoming available as the rail market continues to grow, meaning that an even larger share of the workforce will be relatively inexperienced in coming years (Federal Railroad Administration 2011).

2.2 A New Mode Design

As described previously, safe manual operation of a freight train is a difficult task. An expert engineer develops intuition for handling a train over time, often starting by riding as a conductor alongside an experienced engineer, then during training when they drive with an instructor, and finally through his or her own experience handling the train manually. These experiences teach expert engineers to better anticipate how track conditions, grade, and the particular arrangement and loading of freight cars determines response to brake and throttle inputs. A trainee or engineer that lacks manual control experience may be unable to anticipate train behavior, and the design of new enhanced automation systems must account for this reality.

The goal of the enhanced manual mode is to expand the use of automated modes while preserving the engineer's ability to safely manage the train's movement regardless of the experience level of the engineer (Brooks et al. 2016). Ideally, the new automation mode would allow the engineer to specify the high-level goals or driving strategies of the automation, thereby defining considerations other than fuel consumption that should determine the "ideal" speed profile without requiring the engineer to understand and fine tune various numerical parameters in the speed control system, alleviating much of the need for extensive operator experience. Additionally, the engineer will gain insight into the response of control algorithm since it will create a different optimal profile depending on the particular selection of settings. It is also desirable for the new system to allow the engineer to make changes and compute a new speed profile while the trip is underway in response to unexpected changes, such as a new speed restriction. Such a mode would broadly be defined as a "shared control" model, which generally falls under the category of supervisory control models (Sheridan 2011). For an example of how shared control might work, consider a typical trip along an engineer's route. Before the train departed, the engineer had set some automation parameters that determine a speed plan in line with the operator's goals, like a maximum fuel efficiency or alternatively a minimum trip time. After starting the trip, the engineer gets a call from the dispatcher, who tells the engineer that his train must clear an upcoming intersection by a certain time, or else he or she will have to stop and wait until an approaching train clears the intersection. Viewing the current speed plan, the engineer realizes the train will not be able to clear the intersection, but the engineer is able to adjust the automation parameters to prioritize speed over efficiency until past the intersection, and finds a new plan that follows the new goal. With current in-cab TO automation, which uses a traded control scheme, an engineer is not be able to adjust the speed profile enroute. If the speed plan is too slow to satisfy a required time of arrival at an intersection the engineer has no choice but to take over control manually.

Chapter 3

Problem Statement

To design the new type of automation mode described earlier, it is necessary to understand the operators' normal mental process for controlling the movement of the train. The system must capture an engineer's mental model of train behavior, then understand how this mental model influences behavior through control strategies that the engineer develops. The concept of a mental model is generally defined as a set of internal mental representations of actual systems that are individually constructed based on experience and provide a framework for people to make decisions (Chermack 2003)(Proctor and Van Zandt 2008). These characteristics make it difficult to objectively reconstruct another person's mental model since each engineer will have different experiences and goals during the formation of their models. Nevertheless, it is possible to estimate certain aspects of an expert's mental model and understand the general trends that occur in the mental models of a specific population. These mental model approximations then can be used in various techniques to better understand engineers' control strategies.

3.1 Methods for Analyzing Mental Models

This section briefly describes three methodologies for characterizing the operator's mental model - hierarchical task analysis, cognitive task analysis, and concept mapping - and examples of their prior use in locomotive control research.

3.1.1 Hierarchical and Cognitive Task Analysis

Task analysis as a method can be traced back to the early 1900s, when researchers broke physical tasks into individual "motion elements" that could then be analyzed individually to find efficiency improvements (Proctor and Van Zandt 2006). Both "hierarchical" and "cognitive" task analysis build off of the general idea of defining a top level goal then breaking the tasks necessary to achieve the goal into a hierarchical set of subordinate goals and observable actions that can be separately described in considerable detail, along with precedence constraints. Hierarchical Task Analysis (HTA) generally starts by observing the process or system to be modeled and interviewing the person or people whose mental model is being analyzed about the information and actions they perform. The researcher then determines a series of goals and sub-goals that together represent the task, along with the component physical actions and the plan to achieve each goal or sub-goal (Proctor and Van Zandt 2006). For instance, a hierarchical task analysis for baking a batch of cookies might propose the following goals: procure ingredients, measure ingredients, combine ingredients, form the cookies, bake the cookies. Each of these goals could be further divided into subgoals, such as measuring each specific required ingredient, and would have associated actions, like retrieving a set of measuring cups or a food scale.

Cognitive Task Analysis (CTA), as the name implies, focuses on thought processes as well as actions (Proctor and Van Zandt 2006). This becomes especially relevant as increasing automation shifts work from physical labor to cognitive processes. The key challenge for cognitive task analysis is that the processes of interest are not physically observable, since most of an expert's knowledge is held in his or her procedural knowledge (Clark and Estes 1996). They are only discovered by discussion with a domain expert or inference from observed behavior. As an example, during a hierarchical task analysis, one might observe that a pilot preparing to land an aircraft will adjust the throttle when appropriate, and state that a goal of the pilot is to select the appropriate engine power for landing. A cognitive task analysis would focus on the process by which the pilot decides what engine power to select, including how the pilot assesses current airspeed, altitude, attitude, position relative to the runway, vertical descent rate, and other traffic as well as the cues that indicate to the pilot that he or she needs to add or subtract power. These HTA and CTA techniques have been applied to describe high level aspects of locomotive driving control, as reviewed in later sections.

3.1.2 Concept Mapping

Concept maps are a quick and flexible way to graphically describe relationships among elements in a complex system. This technique helps to externalize and organize knowledge in a variety of settings. Concept maps generally identify main concepts as "nodes" on a graphical network, with lines depicting which nodes influence each other. There are variations in how links are labeled, if at all, and how nodes are organized (Coffey, Hoffman, and Canas 2006). For a complicated task or system without clear divisions of goals or sub-tasks, concept maps lend structure without assuming a pre-existing hierarchy. While task analysis generally has linear progression of goals and subgoals, concept maps can represent more complex relationships between different system elements. Because concept maps do not attempt to discretize processes into independent functions, they are a good choice for systems with many overlapping and interrelated parts. An example concept map can be found in Figure 3-1.

3.2 Prior Work

3.2.1 Branton's Task Analysis, 1978

Though rail transportation has been operational for almost two centuries, it is only in the last 50 years that it has received much attention from human factors researchers. Branton was one of the first researchers to study the human aspects of rail (Branton 1978). Based on interviews with hundreds of engineers and rail inspectors, Branton constructed a theoretical task analysis for train control. He defined each route that an engineer drove as a "mission", and identified three types of task variables, which were climactic, train specific, and geographical. Additionally, Branton argued that there were four essential internal representations of the world: the goal, the operator's position in time and space, the task variables, and the success potential. Furthermore, he hypothesized that these representations serve to help the driver anticipate the train's trajectory and compare it to an "idealized goal". While a formal hierarchy is not presented, the author goes on to outline four informational requirements: the ability to orient oneself, predictive aids, aids for "motivational effort", and immediate environment cues (for example, cues provided by the motion of the train). Branton concluded that if these basic skills were accounted for in new locomotive system designs, engineers would be capable of operating in new environments. This set of proposed internal representations and informational needs forms one of the earliest frameworks for evaluating a locomotive engineer's mental model. However, there is a lack of specific detail on different train handling strategies, only the assertion that such strategies are acquired through experience on the railway.

3.2.2 Naweed et al's Hierarchical Task Analysis, 2018

Currently, two crew members are required for freight rail operations in many countries, also called "two-up" operations. Due to many factors including increased automation capabilities, predicted driver shortages, and potential cost savings, there is some impetus to transition to single-crew operations. In 2018, Naweed et al. analyzed the division of labor between the two crew members using a hierarchical task analysis (Naweed et al. 2018). The intention was to understand the current requirements of a two-person crew in order to determine the feasibility of a one-person crew. The resulting HTA had eight high-level goals, two of which (Driving on the Mainline and Encountering Temporary Speed Restrictions) provide a baseline hypothesis for important elements of the expert mental model being characterized in this project. For instance, the sub-tasks identified in Encountering Temporary Speed Restrictions were used to identify important locations along the simulated route in the experiment for detailed analysis. As with the work completed by Branton, only the goals and sub-goals are discussed, and little to no attention is given to how an engineer and

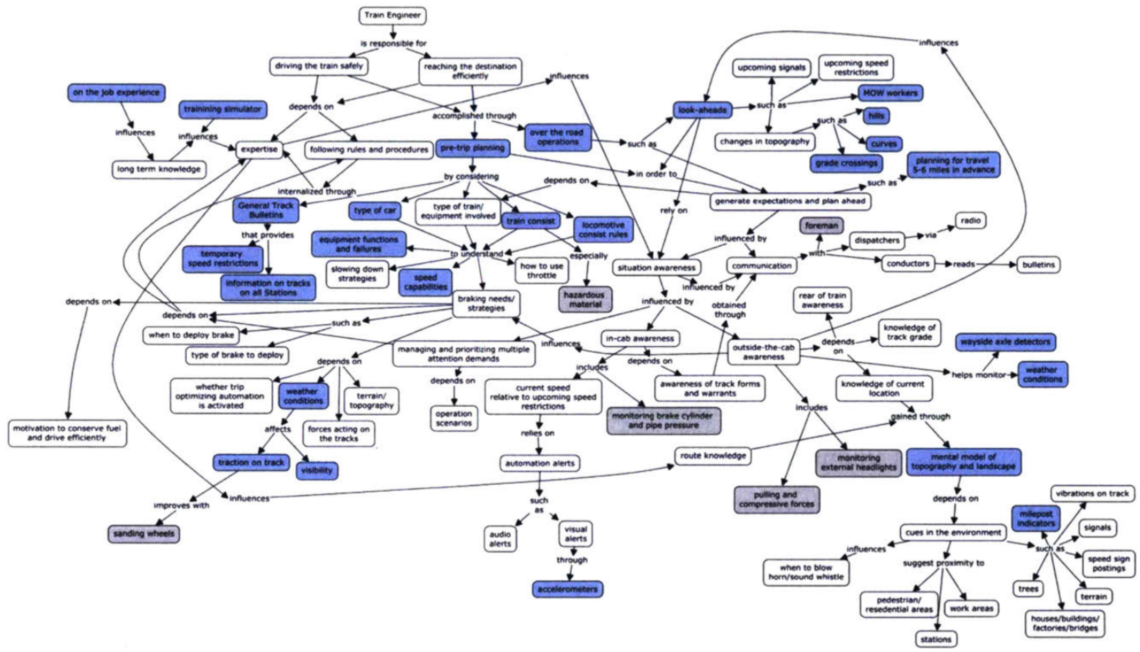


Figure 3-1: A concept map illustrating the duties of a locomotive engineer. Taken from Groshong 2016.

conductor meet these goals (their shared driving strategy).

3.2.3 Groshong's Concept Maps, 2018

In a prior collaboration between GE and MIT, Groshong synthesized conductor and engineer concept maps from three different CTAs previously developed for the FRA (Groshong 2016). One of the final concept maps from this study is shown in Figure 3-1. The intention for this study was to understand the different roles of automation and a human operator at different levels of automation, which then informed paths for further automation in the cab. Like the Naweed HTA, the concept maps also helped identify potential critical areas and cues for engineers that could be analyzed in more detail. Additionally, Groshong identified some key gaps in the existing CTAs from other sources, including specifically a lack of detail regarding braking strategy. The engineer concept map was also used to develop a framework for analyzing verbal interactions during the expert-novice simulations, as is discussed in Chapter 4.

3.3 Key Aspects of the Mental Model for Mode Design

For the purposes of the new automation mode design, these analyses provided helpful background information and informed experiment design, but lacked detailed descriptions for some key aspects of expert mental models. Prior research mainly focused on identifying tasks, information needs, or workload, but did not describe the operators' strategies for accomplishing these tasks or the thought processes of experts. For instance, though the HTA by Naweed breaks the high-level goal of encountering Temporary Speed Restrictions (TSRs) down into a series of sub-tasks that would logically be accomplished in order, it does not pinpoint where the expert engineer begins to factor an upcoming TSR into his or her driving strategy, and simply lists the first sub-task as "observe track sign for upcoming TSR" (Naweed, Balakrishnan, and Dorrian 2018). Since the locomotive engineer is given a "bulletin" that describes the TSRs prior to beginning a trip, it is reasonable to hypothesize that an engineer can plan for an upcoming Temporary Speed Restriction before actually seeing the orange warning placard defining its beginning. Because this cognitive phase of dealing with a TSR does not occur at a pre-specified place, it is not easily captured in a traditional HTA or CTA. Similarly, cognitive maps can help specify important cues, aspects of the environment, and relationships among cues but they do not frame these cues within a spatial or temporal context that explains the decision-making process of engineers. Ultimately, it was decided to represent context by considering the following:

- What the engineer's primary focus is at any given time (including but not limited to goal setting, action planning, information gathering, etc)
- What information triggers an engineer's decision-making process
- Where cues for decisions are located in time and space
- What information is critical to make a decision to act

The goal was to design an experiment that would help to identify these aspects so that they could be incorporated into the design of new automated modes for an improved version of Trip Optimizer.

Chapter 4

Methods

4.1 Overview

As described in Chapter 3, the overall aim of the experiment was to identify the external factors that are part of the engineer's mental model and control strategies when driving a route. Unlike other methods that treat tasks as independent of the overall context, both spatial and temporal positioning are considered to be important for analysis of the engineer's mental model. The study paired an expert freight engineer with a novice subject with very little knowledge of rail operations as the operating crew of a freight train. The subject pair drove two routes together: one with the novice operator controlling the train (NAC, or "novice-at-the-controls") and the other with the expert engineer controlling the train (EAC, or "expert-at-the-controls"). The participant operating the train controls was not able to see the external environment, thus the participants had to verbally communicate the necessary information or instruction to execute an appropriate control action.

From the NAC scenario, it was expected that the expert engineer, through their inquiries and instructions to the novice operator, would reveal the key decisions during driving, what information cued the decision, and how the information was used to arrive at a decision. In the EAC scenario, the expectation was that the content and timing of the information requests from the expert engineer would illuminate both the information engineers rely on for decision-making processes and the frequency at which they updated their mental models. Five expert/novice pairs drove the route, in addition to an expert/novice pair that participated in a pilot study that allowed the research team to refine the study methods. After each experiment, every interaction between the two subjects was coded by the type of interaction and its context. Interactions were then analyzed, along with train handling data, to discern driving strategies and common elements of the mental model and control strategies.

4.2 Subjects

Expert freight engineers were more mostly from a major US or Canadian railroad company (i.e. "Class I railways"). In total, five engineers from five different railroad companies were selected [1]. All had at least six years of freight engineer experience, although some engineers had transitioned to other roles for their railroad and were not driving on a regular schedule. The average age of the engineers was 45.4 years, and the standard deviation was about 10.6 years, with individual subjects ranging between 32 and 61 years old. They had varying levels of familiarity with Trip Optimizer and the rolling map, ranging from one engineer who was qualified as a TO instructor to another engineer who had never used it professionally. None of the engineers were familiar with the route being simulated in the CTIL. Novice subjects were local university students or recent graduates (age 18+) who had no experience with rail operations. Four of the novice subjects were female and remaining subject was male. Both expert and novice subjects were proficient in English.

4.3 Cab Technology Integration Laboratory

The Cab Technology Integration Laboratory, or CTIL, is a full-size, stationary locomotive cab simulator located at the Department of Transportation's Volpe Center in Cambridge, Massachusetts. It is a fixed base training simulator with a full sized freight locomotive replica cab, designed, built and modified for DOT/FRA research

[1] The total number of subjects was limited by available experiment time in the simulator, federal regulations, and scheduling constraints.

use by Alion Science Inc. (McLean, VA) and Corys, Inc. (Corys, Inc., Jacksonville, FL). For this study, it was equipped with standard train handling information displays as well as a Trip Optimizer moving map displays on the conductor's side. The engineer's station was equipped with a 105 side stand for locomotive controls, as seen in Figure 4-1. During all phases of the experiment (excluding training periods), a curtain was put up in the center of the cab, between the engineer's seat on the right and the conductor's seat on the left, and the engineer's forward out-the-window display was blacked out. This allowed the subject seated on the conductor's side to see out of the cab as normal, while the subject seated at the engineer's control position could not see the outside environment ahead, nor the moving map on the conductor's side.

Figure 4-1: The right side of the Cab Technology Integration Laboratory, where the engineer typically sits. The control stand can be seen to the left of the seat.

Figure 4-2: The left side of the Cab Technology Integration Laboratory, where the conductor typically sits.

4.4 Test Conditions

4.4.1 Overall Study Design

In order to understand different facets of expert mental models, two test conditions, referred to as "scenarios", were designed. Both scenarios shared the same route, which was a 65-mile segment of the BNSF Railway's Aurora subdivision in Illinois that was generally flat with some rolling hills outside urban areas. The scenarios differed in the roles of each of the test subjects, as well as the placement of different features along the route, such as speed restrictions and yellow or red signals. The scenarios were designed to require the expert subjects to make explicit train control decisions that would be verbally communicated and explained, thus providing some insight to the underlying mental model. Some scenario events were known well before being encountered, such as a temporary speed restriction, since they were outlined in the simulated track bulletins or in incoming dispatcher messages. Other events such as signal states were manipulated along the route (e.g., changed from green to yellow) and could only be acted upon when observed. Still other events dealt with timing and issues relating to system-wide traffic flow. Staying on schedule is important as any delays in the scheduled arrival may complicate the schedule of the train across the network. It was assumed that additional time pressure might alter the engineer's driving strategy and would reveal how their decision-making process would alter the content or timing of information when controlling the train.

The scenarios were named according to the subject in charge of manipulating the train controls. Both the novice-at-the-controls (NAC) and the engineer-at-the-controls (EAC) scenarios were designed with five common Key Decision Points ("KDP"s) at which behavior could be compared. The KDPs included three temporary speed restrictions and two red signal encounters requiring a full stop. The KDPs were the primary points around which it was hypothesized that the expert subject would need to make explicit control decisions. The locations of these KDPs varied between the two scenarios to avoid any effect of participants learning and anticipating the events, but all of the subject pairs drove the same NAC and EAC scenarios.

In addition to the five KDPs, one timing-related even was included. During the experiment, the dispatcher (an experimenter) notified the train crew of a new "Meet-and-Pass" about 10 miles before a selected siding. The dispatcher informed the expert that another train was waiting in a siding ahead, and that the crew of the stopped train was about to reach the end of their duty day (at which point the crew would legally be required to stop operating the train). The dispatcher also asked for the expert subject's estimated time to the siding. The scenario was designed to create a sense of urgency in train operations and to determine if time pressures help inform decision making en route. A full description of each route with locations of each event is in Appendix A.

4.4.2 Novice-at-the-Controls Scenario

In the NAC scenario, the expert instructed the novice as they drove the route together, explaining how to manipulate the controls and what factors had led to a specific decision. Presumably, the novice possessed only a rudimentary mental model of train operations and basic control strategies based on their brief pre-experiment training, so the expert subject would need to explicitly communicate all of the different aspects of their mental model to the novice to drive the trip safely. The communication guidelines for the NAC scenario instructed the expert subject to give directions and a brief explanation for those directions, while the novice subject was to ask questions whenever directions were unclear. The expert was also responsible for the safe operation of the train, and to issue instructions to the novice that mimicked how the expert would drive if at the controls. The expert could consult a paper track chart and any speed bulletins throughout the scenario. The expert subject, who was seated at the conductor's console, also had access to basic train handling data, including speed, acceleration, notch position, and brake pipe pressure, as well as the Trip Optimizer moving map, which showed location, track speed (not including temporary speed restrictions), and grade. The novice subject had no view of the external environment and only access to the standard instrument displays in the locomotive cab such as train speed, air brake pressures, and acceleration, but not the Trip Optimizer moving map.

4.4.3 Engineer-at-the-Controls Scenario

In the EAC scenario, the expert controlled the locomotive, but could not see out of the simulated cab and could not see the TO moving map display on the conductor's side. Experts were given a paper track chart with grade information removed and were able to view basic train handling data. For the purposes of this experiment, basic train handling data included speed, acceleration, brake pipe and reservoir pressure information, throttle and brake lever positions, and distance measuring device readings. The novice subject sat in the conductor's seat on the left side of the cab and was able to see out the external environment. In addition to standard train displays, the novice subject could see the TO moving map. This objective of this scenario design was to study what external information was essential to constructing and updating the expert's mental model. The expert would presumably need to constantly request location and train state information from the novice, and the content and frequency of these requests would reveal the information the expert needed but lacked. In order to prevent the interactions being dominated by certain requests for information that would certainly be needed for safe locomotive operations, the novice was permitted to proactively report seeing a milepost, rail sign, or rail signal. Additionally, expert subjects could give novices standing orders to report specific conditions, such as reporting when the end of the train crested the top of a hill.

4.4.4 Condition Summary Table

Table 4.1 summarizes the different test conditions. The subject seated on the right, which is the engineer's seat in normal operations, was able to manipulate the train controls from their position. The "signal reference guide" was a paper chart describing types and meanings of signals that could be encountered along the route. It was given to both expert and novice subjects.

4.5 Protocol

4.5.1 Informed Consent

The experimental protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES). Informed consent was obtained from both novice and expert subjects before beginning any portion of the experiment.

4.5.2 Training

No more than seven days before the experiment, the novice signed the informed consent form, and then completed a short training session at the CTIL facility. The

Scenario Type	Conductor's seat (left)	Engineer's seat (right)	Expert Subject Resources	Novice Subject Resources	No. of Speed Restrictions	No. of Red Signals	No. of MP Interactions
NAC	Expert	Novice	• Basic train handling information • TO moving map • Track chart, with grade • Signal ref. guide • Track bulletins • Trip rules • Communication guidelines	• Basic train handling information • Signal ref. guide • Communication guidelines	3	2	1
EAC	Novice	Expert	• Basic train handling information • Track chart w/o grade • Signal ref. guide • Track bulletins • Track bulletins • Trip rules • Communication guidelines	• Basic train handling information • TO moving map • Signal ref. guide • Communication guidelines	3	2	1

Table 4.1: Comparison of the engineer-at-the-controls and the novice-at-the-controls scenarios. In the last column, "MP Interactions" refers to meet-and-pass interactions.

novices watched a brief slideshow overview describing the CTIL simulator, the primary cab controls that would be used (notch, dynamic brake, and air brake), and the basics of rail signaling and signage. Then they moved into the CTIL cab for a ten minute trial drive, during which operation of throttle, dynamic brake, and air brake, alerter, and TO moving map were explained. The novice drove along the route for a short distance, maintaining the maximum track speed. Signals and signage along the track were pointed out to the novice when appropriate. The novice completed the training session by bringing the train to a full stop at a red signal set by the experimenter.

On the day of the experiment, the expert subject arrived before the novice subject to complete the expert training session. After providing informed consent, the expert viewed a brief slideshow detailing the train and its properties (weight, length, load distribution) as well as the details of the route, including its length, track speed, and standard signals. The expert subject was then given generic "trip rules" (Appendix C) and a track chart, and allowed as much time as needed (typically five to ten minutes) to familiarize himself with the track chart for the route. When the expert subject felt prepared, they moved to the CTIL cab where they were given a brief overview of the CTIL, and then drove a training route. The training route was over the same route that the actual experiment took place over. The main difference between the training scenario and experimental scenarios was the lack of Key Decision Points – there were no temporary speed restrictions and no interactions with the dispatcher.

4.5.3 Experiment Session

After successful completion of the training route, the expert was introduced to the novice and both subjects were briefed on the two experimental scenarios, EAC and NAC. One scenario was performed before lunch, and the other after lunch. Three of the subject pairs performed the NAC scenario first and the other two subject pairs started the EAC scenario first. Subjects were briefed on applicable track bulletins and communication guidelines for the scenario. The track bulletins could be used during the trial, as was a reference sheet for the signals for both novice and expert.

Since wayside signalling indication schemes are not standardized across all railroads, some expert subjects felt more comfortable with a reference sheet showing what the wayside signals meant. A track chart that had been stripped of grade information was provided to the expert for the EAC scenario; for the NAC scenario the expert sitting in the conductor's sea was allowed to reference the regular track chart that included grade data. Each scenario took about 75 minutes to complete. After completing the first route, debrief questionnaires were given to both expert and novice subjects asking both subjects for their general impressions of how the scenario went, both in terms of the train handling and in terms of interactions. This was primarily to highlight any instances from the trial that would merit closer scrutiny from the research team, for instance, if expert subject had felt, at any point, that the train was not being controlled in a safe manner. A researcher also led a short verbal debrief to follow up on any points of interest from the trial. For example, one expert subject asked very few questions of the novice during the EAC scenario, and the researcher followed up by asking if the expert subject was aware that they could ask for more information throughout the scenario. Lunch followed the first scenario. Briefings for the second scenario followed the same pattern as the first trial. As before, subjects completed a questionnaire when the route was finished and were again verbally debriefed. Subjects were given an opportunity to ask any remaining questions and were given contact information for the research team in case they required any follow-up.

4.6 Data Collection and Processing

4.6.1 Data Collection

There were two primary sources of data: an audio/video recording of the interior of the cab, and the extensive train handling data recorded by the simulator. The audio/video recording also captured the engineer's (right seat) primary internal display with the basic train handling data as well as the simulated view outside of the cab. The dialogue of the interactions between expert and novice were captured in the audio stream and subsequently analyzed as described below. The digital train handling data recorded 34 specific fields. For the purposes of this experiment, the most important data fields were: time, distance, grade, speed, acceleration, brake pipe pressure, and notch (throttle setting).

4.6.2 Interaction Coding

To anonymize the data, subject pairs were assigned eight-character pseudorandom codes. These codes were then truncated to the first two letters for ease of use. The following subject codes appear throughout the Results and Discussion sections: BH, CP, HD, NW, and WW.

After obtaining the interactions between the novice and the expert subjects, a special purpose ad-hoc coding scheme was developed and used to categorize the interactions between the expert and novice by content, location and timing, with the intent of capturing the underlying mental model and decision-making process of an engineer. The coding framework was largely based on the techniques used in computational linguistics, with domain-specific knowledge from concept maps developed in previous locomotive studies (Groshong 2016). Interactions were first given a surface-level (Level 1) code, which expressed the type of interaction (question, response, acknowledgement, confirmation, etc). Level 1 codes encompass Forward and Backward Communicative Functions as outlined by Core and Allen (Core and Allen 1997). The interactions were given a second code related to the semantic meaning of the interaction (Level 2) code. The categories for Level 2 were largely influenced by what interactions were anticipated based on previous work. A more complete description of how to use each Level 1 and Level 2 code is also available in Appendix F. Table 4.2 summarizes the applications of each of these codes, along with the 3-letter reference label that was used during the coding process.

Additionally several subcodes, listed in Table 4.3, were used to provide additional detailed information on the content of the interaction, not just the semantic purpose. For example, an interaction where the expert subject asks about upcoming grade would have a query-wh Level 1 code (QWH), a check precondition Level 2 code

(CHK), and a "grade" subcode. Two coders, both MIT undergraduate students with no prior knowledge of freight operations, were first trained on the basics of railway operations, including driving the simulator, then trained on how to apply the coding scheme as they watched the video and audio recordings. During the training process, the coders' results were checked against an example training set of interactions from the test run of the study to ensure that the coding would be applied as intended.

Complete transcripts of the video recordings for each trial were created with each verbalization recorded with a timestamp. Important events, such as a passing a milepost, were also recorded. The transcription process was spread out over multiple members of the research team. These transcriptions were then given to the two coders who analyzed each trial individually, producing two different encodings for the same set of interactions. After all trials were coded, a single "reconciled" encoding was produced for each trial by merging the two encodings, which was then used for the analysis. Conflicts between the individual encodings were resolved through discussion on the part of the coders, with each coder explaining his or her reasoning for a code, and then the two coders agreed on the most reasonable explanation. Sections of each trial were also selected for quality assurance tests, where a member of the research team would code that portion of interactions and compare their conclusions with the reconciled encoding for consistency.

Level 1 Codes		
Acknowledge	ACK	A verbalization indicating acknowledgement of the previous statement or request
Confirm	CON	A verbalization indicating the person agrees with the previous statement or request
Dispatcher	DIS	A code used to indicate interaction with the dispatcher, which could be either the dispatcher's verbalizations or responses from the subjects to the dispatcher
Inform	INF	Any verbalization that communicates information, while not being an inquiry
Propose	PRO	A verbalization that suggests a choice to either follow or not follow a request
Query-if	QIF	A conditional question, typically can be answered with a yes or no
Query-wh	QWH	A question concerning who, what, when, where, why, how
Request	REQ	A statement that indicates the action to be taken by the addressee
Reject - explicit	REX	A verbalization indicating the person disagrees with the previous statement or request, and refutes the previous statement or request directly
Reject - implicit	RIM	A verbalization indicating the person disagrees with the previous statement or request, and indirectly refutes the previous statement or request
Level 2 Codes		
Check precondition	CKP	Collection of information on preconditions (environmental cues, information from paperwork) to prepare for an upcoming scenario
Remind plan	RMD	Any reminders of pre-trip planned events
Confirm plan	CNF	A verbalization that confirms the upcoming scenario in response to a reminder or discussion of preconditions
Evaluate plan	EVL	A verbalization that reconfirms what should be done within the specific context of the scenario
Critique plan	CTQ	A verbalization that critiques or questions the rationale or execution of a plan
Refine plan	REF	Expanding on a previous plan to add new details or make changes based on new information or circumstances
Clarify plan	CLF	A question or clarification to ensure plan execution conforms to expectations
Explain plan	EXP	The rationale and reason for executing a plan a certain way
Execute plan	EXC	The action or set of actions that should be employed to execute a plan

Table 4.2: An explanation of each Level 1 code and Level 2 code used during the coding process, along with the unique three-letter abbreviation for each code.

Controls	**Displays**	**Train**	**Train**
• Air brake • Dynamic brake • Throttle • Notch	• TO display • Accelerometer • Counter/Distance Measuring Device (DMD)	• Turn out • Curve • Grades • Switch • MP *Y* • Speed flags • Speed restrictions • *X* MPH (track speed limit)	• Train type • Slack action • Track forces • Train breakage • Front, middle, or end of train • Train speed change

Table 4.3: A table of all subcodes used. The subcodes were sorted into different categories for easy reference, but only the subcodes were applied during the coding process (there is no distinction between subcodes of different categories in the encoded transcript).

Chapter 5

Results

5.1 Aggregate Data Analysis

5.1.1 Novice-at-the-Controls Scenario

The novice-at-the-controls scenario was designed to reveal the context of the engineer's decisions made along the route. This was accomplished by asking the expert engineer to verbally communicate their higher-order thinking processes to the novice subject through their instructions and the reasoning behind those instructions. As a result, the analysis for the NAC scenario focused on the Level 2 interaction types which identified the different types of higher-level processes, rather than the form of the interaction, which was represented by Level 1 interactions. Figure 5-1 shows the frequency of both the Level 1 (top) and Level 2 (bottom) interactions.

For Level 1 codes, 90% of the total interaction types were "request" (REQ), "acknowledge" (ACK), and "inform" (INF), which made up 41%, 25%, and 24% of all interactions, respectively. This is not particularly surprising since the NAC scenario primarily consists of the expert subject instructing and giving background information, which are "request" and "inform" interactions, and the novice subject complying with the request, which was verbalized with an "acknowledge" interaction. The least common interactions were the rejection codes (REX and RIM) and "propose" (PRO), each of which made up less than 1% of all interactions. Again, this is an understand-

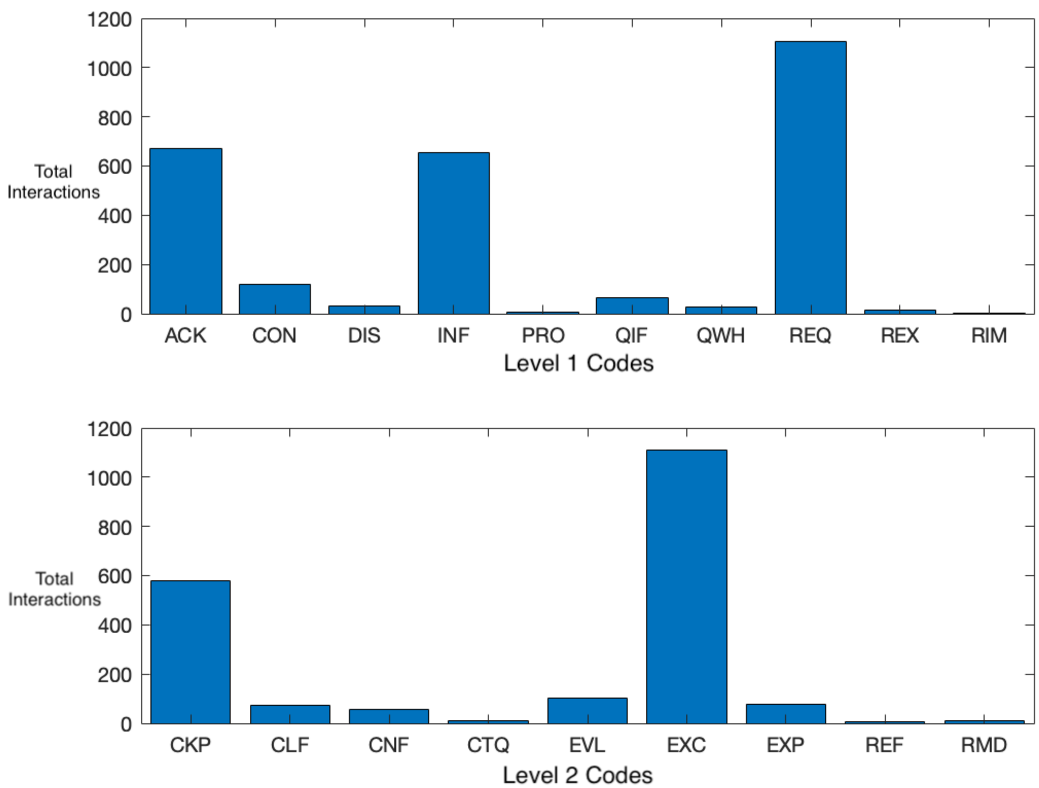


Figure 5-1: Frequency of all novice-at-the-controls scenario interactions. The top plot shows the Level 1 codes, the bottom plot shows the Level 2 codes. Note that two interaction types, "EXC" or execute and "CKP" or check precondition, dominate the Level 2 interactions.

able result as the novice subject, having little knowledge of rail operations, is not likely to reject actions requested by the engineer nor propose his or her own action plan. In turn, the expert is unlikely to have a proposal to reject, and generally instructions from the expert would be "request", not "propose", so these codes were not very applicable in this scenario.

For Level 2 codes, the most common interactions were "check precondition" (CKP) at 28% and "execute" (EXC) at 55% of all interactions with a Level 2 label[1]. The frequency of "execute" is unsurprising, as almost any manipulation of the cab controls by the novice would have been prompted by an "execute" interaction from the expert. The "check precondition" interaction generally indicates an assessment of the current situation for any cues that would trigger a pre-defined action sequence. The key idea is that the mental task at that moment, as represented by the interaction, is primarily one of maintaining situational awareness. The dominance of the "check precondition" interaction over planning-focused interactions indicates that the expert's primary task enroute was looking for situational cues that might subsequently trigger strategic or tactical planning. The high frequency of the "check precondition" interaction lends some insight into an expert's mental model that will be discussed in depth in Section 6.

To reveal the distribution of the interaction types, histograms (e.g. Figure 5-2) were generated for each Level 1 and Level 2 code along the entire route, grouped in 0.25 mile increments. Plotting by distance means that interactions that occur when the train is moving more slowly will be condensed into fewer bins compared to interactions that occur when the train is moving at high speed. However, there was no alternative since the engineers drove the route with different speed profiles, histograms referenced to the time elapsed would not have a one-to-one relationship with position of the train along the track.

The peaks in the Level 2 interaction histograms (Figure 5-2) at milepost 95 and

[1]Since many interactions were a simple acknowledgment on the part of either the novice or the expert subject, there are numerous interactions that do not have a Level 2 code assigned. Interactions without a Level 2 code made up approximately 25% of all interactions during the NAC scenarios. When calculating the frequency of appearance of different Level 2 codes, the percentage is taken out of interactions that have been labelled with a Level 2 code.

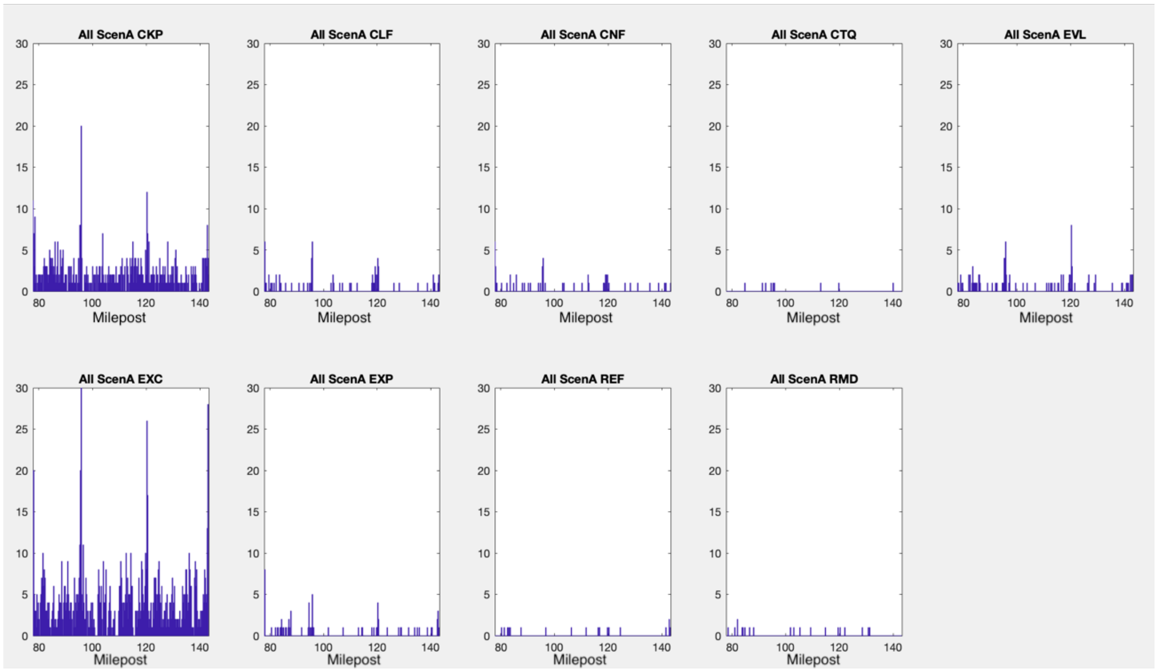


Figure 5-2: All Level 2 novice-at-the-controls scenario interactions binned according to their code and the milepost at which they occurred. The spikes around milepost 95 and again around milepost 120 are stop signals along the track.

milepost 120 coincide with the locations of stop signals along the track, which were Key Decision Points 2 and 4. For reference, the key decision points for the NAC scenario are listed in Table 5.1 (below). The temporary speed restrictions, which were Key Decision Points 1, 3, and 5 (as described in Section 4.1), did not appear to cause a significant spike in interactions of any type .These occurred from mileposts 87 to 88, 113 to 114, and 130 to 131. It should be noted that there is a compression effect at lower speeds (such as around stop signals) because the interactions are plotted over distance, rather than time. As such, the spikes in themselves are not necessarily indicative of a higher interaction frequency. The least frequent Level 2 interactions were "critique plan" (CTQ), "refine plan" (REF), and "remind plan" (RMD), each making up 1% or less of all interactions with a Level 2 label. This indicates that plans were generally not being adjusted enroute, and that neither the expert nor the novice subject felt it necessary to remind the other subject of the plan.

To further investigate use of the "check precondition" and "execute" interactions, histograms for these codes were more closely analyzed for a five-mile-long section

KDP #	Type	Milepost
1	1-mile temporary speed restriction	MP 87-88
2	Approach signal (yellow) followed by stop signal (red)	Yellow: MP 93.9; Red: MP 95.9
3	1-mile temporary speed restriction	MP 113-114
4	Approach signal (yellow) followed by stop signal (red)	Yellow: MP 118.6; Red: MP 120.6
5	1-mile temporary speed restriction	MP 130-131

Table 5.1: A description of each key decision point for the NAC scenario and where it occurred along the route.

.

of the route around Key Decision Point 5 (a one mile TSR at MP 130). Looking at the "check precondition" histogram centered around the MP 130 temporary speed restriction in Figures 5-3, areas of increased interaction can be linked to critical points along the route. First, the peak two miles before the temporary speed restriction (MP 128) reflects the activity occurring when the crew encounters the speed restriction warning flag [2] along the track. This warning flag represents a precondition that must be acted upon by the crew. The next peak occurs at the end of the TSR (MP 131), where a "resume speed" flag informs the engineer they are leaving the speed restriction. Finally, a peak is seen when the rear end of the train exits the speed restriction and the engineer can resume track speed. Again, the end of the TSR is a precondition that, when met, will trigger a series of pre-determined actions by the crew. For the same 5-mile section, the main peak of the "execute histogram (Figure 5-4) occurs about a quarter mile before the TSR, indicating that the engineer was making the most adjustments during this period. Together these plots help to differentiate where an engineer might be concerned with making a decision (the areas of high "check precondition" activity) from where an engineer is primarily concerned with acting on a decision (areas of high "execute" activity).

[2] A speed-restricted area in this experiment was marked by two "flags", which are colored placards placed along the track. A yellow flag marked the start of the speed restricted area, and a green flag marked the end (though it should be noted the train needed to maintain restricted speed until the end of the train had passed the green flag). Additionally, a yellow flag was placed two miles before the start of the speed restricted area to warn or remind the subjects of the upcoming speed restriction.

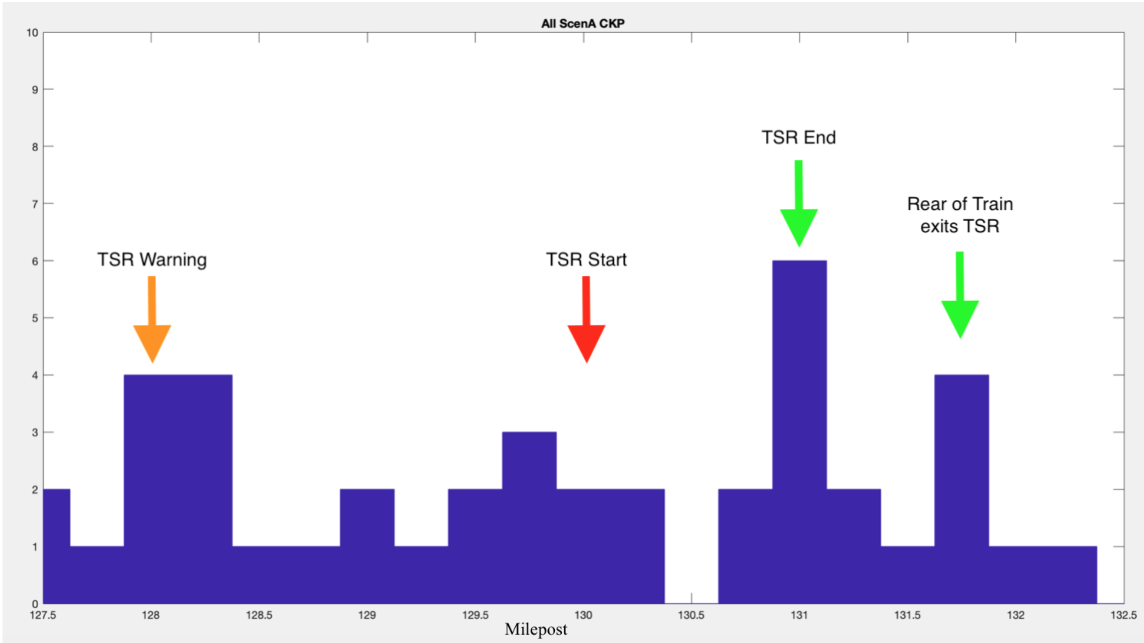


Figure 5-3: Frequency of the "check precondition" interaction in the novice-at-the-controls scenario, over a five-mile section of the route around Key Decision Point 5, which was a temporary speed restriction.

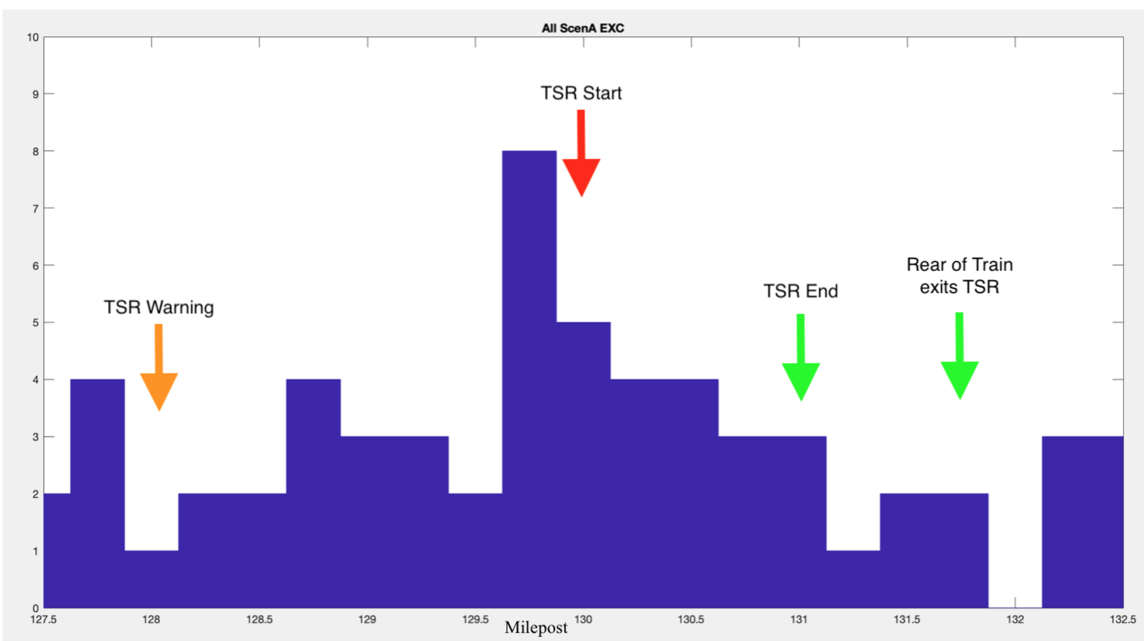


Figure 5-4: Frequency of the "execute" interaction in the novice-at-the-controls scenario, over a five-mile section of the route around Key Decision Point 5, which was a temporary speed restriction.

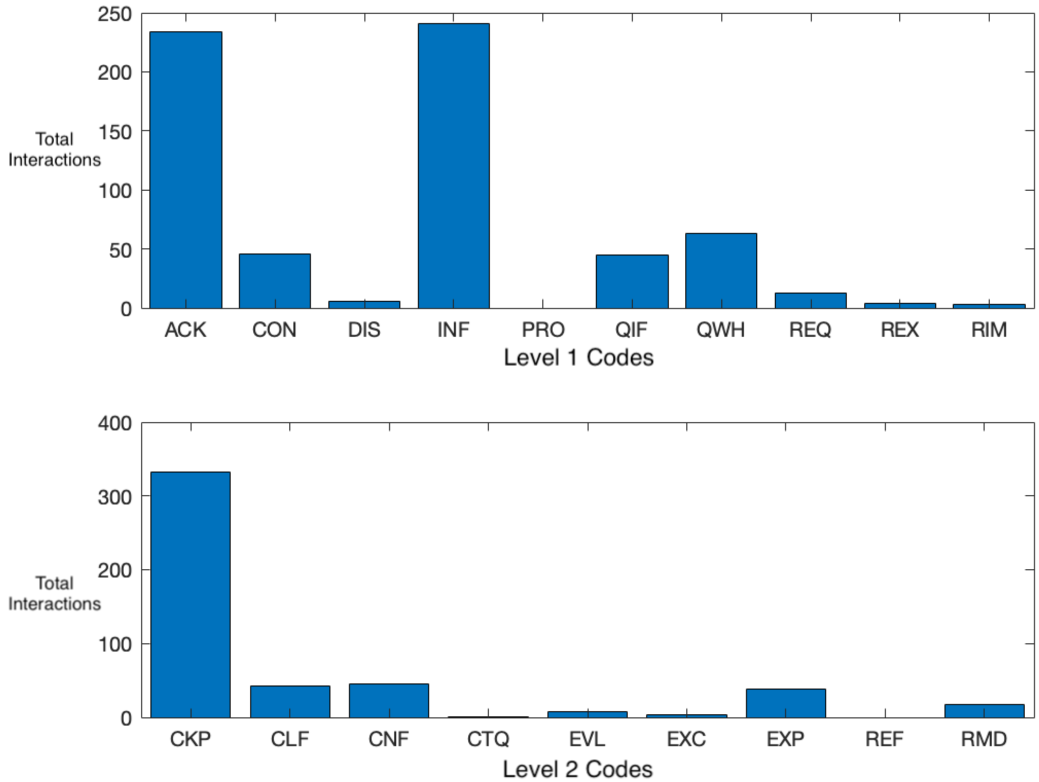


Figure 5-5: Frequency of all engineer-at-the-controls scenario interactions. The top plot shows the Level 1 codes, the bottom plot shows the Level 2 codes.

5.1.2 Engineer-at-the-Controls Scenario

The engineer-at-the-controls scenario was designed to highlight the information that an expert engineer uses in his or her decision-making process. The most common Level 1 codes from the EAC scenarios were "inform" (INF) at 37% of all interactions and "acknowledge" (ACK) at 36% of all interactions. These interactions likely represented information being passed from the novice to the expert subject and the subsequent acknowledgement that the information was received. The two query codes, query-if (QIF, usually requiring only a yes/no answer) and query-wh (QWH, usually a what, when, where, why open-ended question), made up another 15% of interactions, with query-wh being slightly more common than query-if.

The query codes were less common than originally hypothesized. There were just over 100 combined query interactions for all five subject pairs, so each subject averaged 20 queries. Given that the route driven was over 60 miles long, this works out to an average query rate of less than once every three miles. The dearth of queries is partially explained by the much higher frequency of the "inform" code. Since novice subjects were often instructed by the expert to preemptively inform the engineer of mileposts, signals, and signs, the expert subject did not have to actively query the novice. Before conducting the experiment, it was hypothesized that the frequency of query codes would serve as a proxy for the frequency with which an expert subject updated his or her mental model. However, analysis of these results shows that the expert subject had too much access to external information, most notably the modified track chart, to use the query codes frequency as a heuristic.

Finally, the least common Level 1 codes were "propose" (PRO), which did not appear, and both rejection codes (REX and RIM), which combined yielded 1% of interactions. Again, the lack of propose interactions is unsurprising, since there would be little need for the novice to propose an action, since expert subjects were relying on novice subjects only for tactical information (the location of the train, outside conditions) and not strategic advice. Likewise, experts rarely propose plans, as instruction from the expert subject would generally be coded as "request".

For Level 2 codes, "check precondition" (CKP) was by far the most common code, at 68% of all interactions with a Level 2 label [3]. The next most frequent codes were "clarify plan" (CLF), "confirm plan" (CNF), and "explain plan" (EXP), with 9%, 9%, and 8% of Level 2 codes, respectively. As with the NAC scenario, the high frequency of "check precondition" codes indicated that the expert subject was generally following a predetermined plan and most interactions were intended to assess a situation. Once established, the plan might be clarified, confirmed, or explained, but never refined in this scenario - "refine plan", or REF, did not appear in any of the EAC interactions. Uncommon codes were "critique plan" (CTQ) and "execute" (EXC), which were each less than 1% of Level 2 interactions. The only instance of a CTQ was initiated by the expert who was driving, not the novice. The lack of

[3] As before, the Level 2 frequencies were calculated out of the total number of interactions with a Level 2 code applied. Interactions without a Level 2 code made up roughly 25% of all interactions, which was similar to the NAC scenario.

"critique plan" is consistent with the absence of "refine plan", since these codes would be expected to occur together, e.g., a critique of the plan would trigger a refinement of the plan. The low frequency of "execute" interactions reflects the fact that the EAC scenario did not require the expert subject to give many instructions to the novice, just occasional standing orders. Typical standing orders included informing the expert each time a milepost was passed, or when a signal came into view.

Histograms were generated for each Level 1 and Level 2 code across the entire route, again grouped in 0.25 mile increments, shown in Figures 5-6 and 5-7. The stop signals along the route at milepost 105 and milepost 122 are identifiable from the peaks in interaction density for both Level 1 and Level 2 codes, which was also the case with the NAC scenario. As in the NAC scenario, route events, like temporary speed restrictions, did not appear to cause easily identifiable higher interaction densities. Key Decision Point locations for this scenario can be found in Table 5.2.

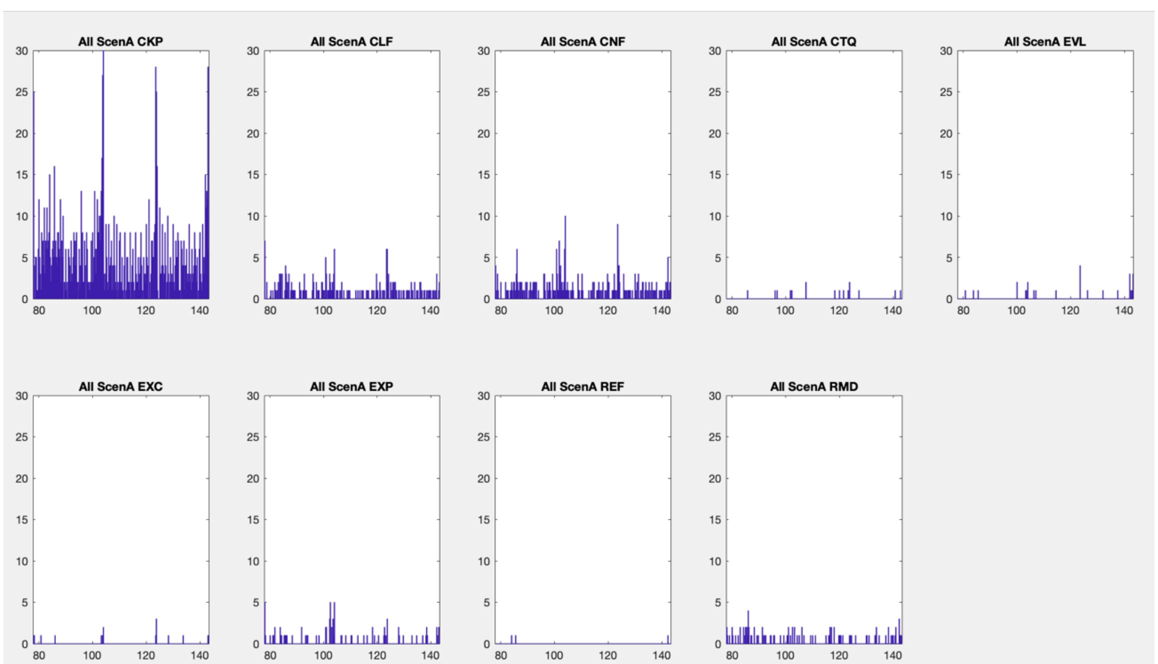


Figure 5-6: All Level 2 engineer-at-the-controls scenario interactions plotted against the milepost at which they occurred. The spikes around milepost 105 and again around milepost 122 are stop signals along the track.

KDP #	Type	Milepost
1	1-mile temporary speed restriction	MP 80-81
2	Approach signal (yellow) followed by stop signal (red)	Yellow: MP 101; Red: MP 104
3	1-mile temporary speed restriction	MP 118-119
4	Approach signal (yellow) followed by stop signal (red)	Yellow: MP 122.4; Red: MP 123.9
5	1-mile temporary speed restriction	MP 132-133

Table 5.2: A description of each key decision point for the EAC scenario and where it occurred along the route.

.

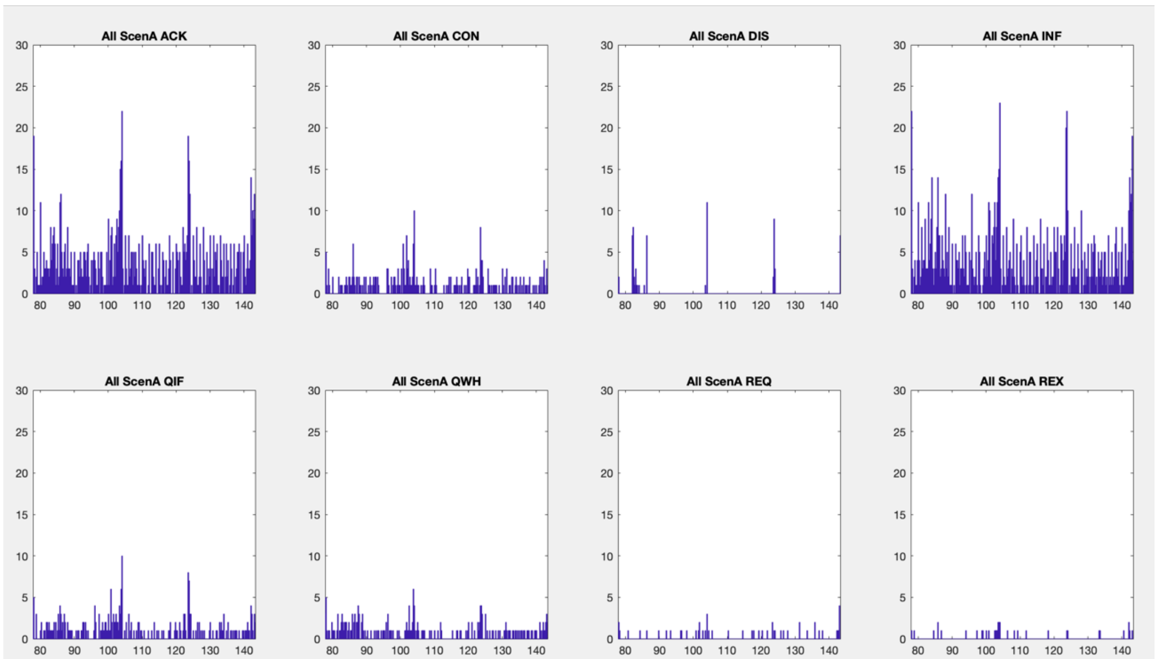


Figure 5-7: All Level 1 engineer-at-the-controls scenario interactions plotted against the milepost at which they occurred. The spikes around milepost 105 and again around milepost 122 are stop signals along the track.

Though the query codes were not analyzed as originally intended, their infrequent occurrence does imply that the expert subjects generally did not require very frequent updates of information from outside the cab environment. Prior to beginning the trip, some expert subjects gave standing orders to the novices to report mileposts, track signals, and track signage when first seen. Qualitatively, it appears that these real-time updates combined with the information provided to them before the simulated trip (all TSR locations and a track chart with grade information removed) was sufficient to safely drive the train for most of the route. Interestingly, one expert subject rarely queried the novice during the trial, but also significantly exceeded the speed limit multiple times during the trial. This example suggests that there may be a minimum frequency for updating information about the outside environment for safe operation of a train, even if those updates are relatively rare.

The content of the query codes was captured by the use of subcodes, which identified the set of wayside objects, such as signals, or physical characteristics affecting train state, such as grade, that were the subject of the query. As seen in Figures 5-8 and 5-9, the most common subcodes for both types of queries were “milepost”, “signal”, “grade”, and “speed restriction”. This indicates that these environment objects or attributes are the most important for the engineer to update in real-time. Interestingly, the proportion of each of these subcodes varied between the two types of queries. For example, the most common subcode for query-if was “signal”, but the “signal” subcode was much less common for the query-wh subcodes, indicating that queries about signals were more likely to be answerable with a yes or no than an open-ended question. In other words, questions like “can you see an upcoming signal?” or “is the signal green?” were more likely than “what color is the upcoming signal?”. On the other hand, mileposts were by far the most common query-wh subcode, indicating a preference for questions like “what milepost are we at?” vs questions like “are we passing milepost 101?”.

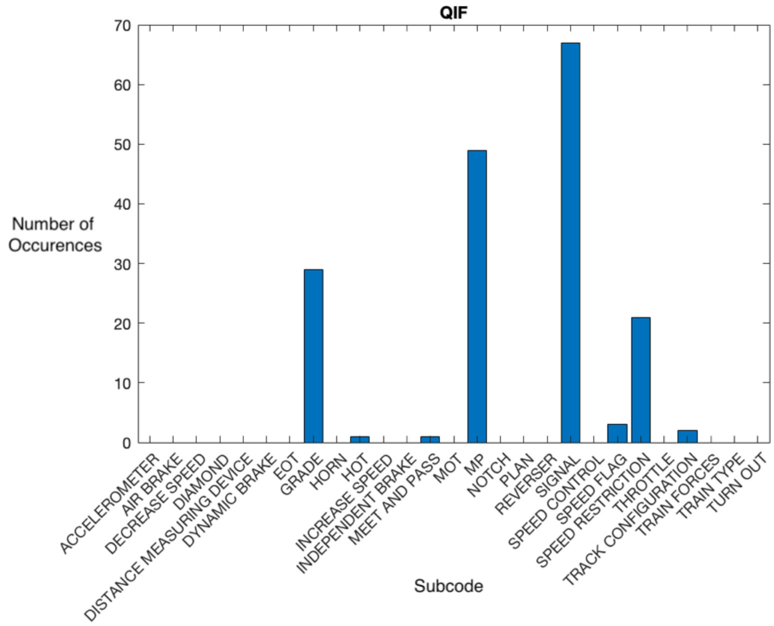


Figure 5-8: Frequency of all subcodes over all engineer-at-the-controls scenarios for the query-if interaction code. The -if in query-if denotes that the question could be answered with a yes or no, which distinguishes it from the other query code, query-wh.

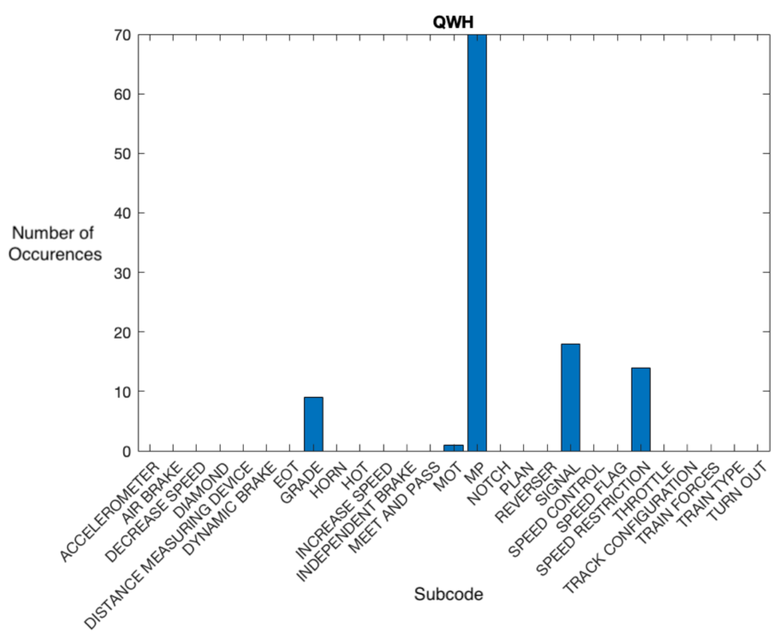


Figure 5-9: Frequency of all subcodes over all engineer-at-the-controls scenarios for the query-wh interaction code. The -wh in query-wh stands for who/what/when/where/why, and this code was used for any question that could not be answered with a yes or no.

5.2 Train Handling Differences Between Expert Engineers

In addition to an analysis of all interactions from all subject pairs, results from each subject pair were compared to determine similarities and differences between expert subjects' control strategies. To understand the relationship between the decisions made along the route and the interactions between the crew, each interaction, train speed, and notch was plotted by its location along the route. Over short intervals near Key Decision Points, it is possible to compare train handling behavior between engineers and to understand aspects of each engineer's driving strategy.

Comparison plots were generated for each engineer around each of the five key decision points (KDP), three temporary speed restrictions and two stop signals. Two sets of these comparison plots, both from the NAC scenario, are presented and analyzed in this section. The first set of plots is from KDP 4, which is a yellow approach (slow down) signal followed by a red stop signal. The second set of plots is from KDP 5, which is a temporary speed restriction. These KDPs were chosen for deeper analysis because they had the least potential to be influenced by other events along the route (permanent speed restrictions, the meet and pass interaction, etc).

For each subject, the top plot shows the actual speed vs speed limit with interactions marked by various symbols, the middle plot shows the throttle/dynamic brake setting, and the bottom plot shows brake pipe pressure. Throttle settings are discrete while the dynamic brake can be adjusted through a continuous range which has been scaled in these figures to provide better detail. On the bottom plot, a decrease in brake pipe pressure denotes that train air brakes have been applied, with larger decreases indicating higher braking force.

5.2.1 Key Decision Point 4, NAC Scenario

In Figures 5-10 to 5-14, the red line represents the speed limit at the location while the blue line shows the actual train speed. It should be noted that at MP 118.6, the speed limit appears to drop to 40mph. In reality, there was a yellow signal at MP 118.6, which indicates that the train should begin to slow to 30 mph in preparation to stop at the next signal. Thus, from MP 118.6 to MP 120.6, the speed limit was not strictly 40 mph, which is why all of the subjects appear to be exceeding the speed limit in the initial part of this segment. The "execute" interactions are marked with a black star and indicate where the expert chose to begin executing the current task. The "check precondition" interaction is marked with a magenta plus and indicates where the subject pair is evaluating a situation to determine a set of appropriate actions. Thus, the location and timing of these actions can be compared for each of the different subjects.

Comparing these plots reveals a number of different preferences for slowing the train to a stop. First, engineers varied significantly in their speed profiles over the two miles between the first signal (yellow) and the second signal (red). As discussed above, the yellow signal at MP 118.6 nominally indicates engineers should slow to about 30 mph. Some subject pairs (HD and CP) began to slow almost immediately after passing this signal, while others took a bit longer (NW and WW). Subject pair BH did not slow down appreciably until more than a mile past the signal.

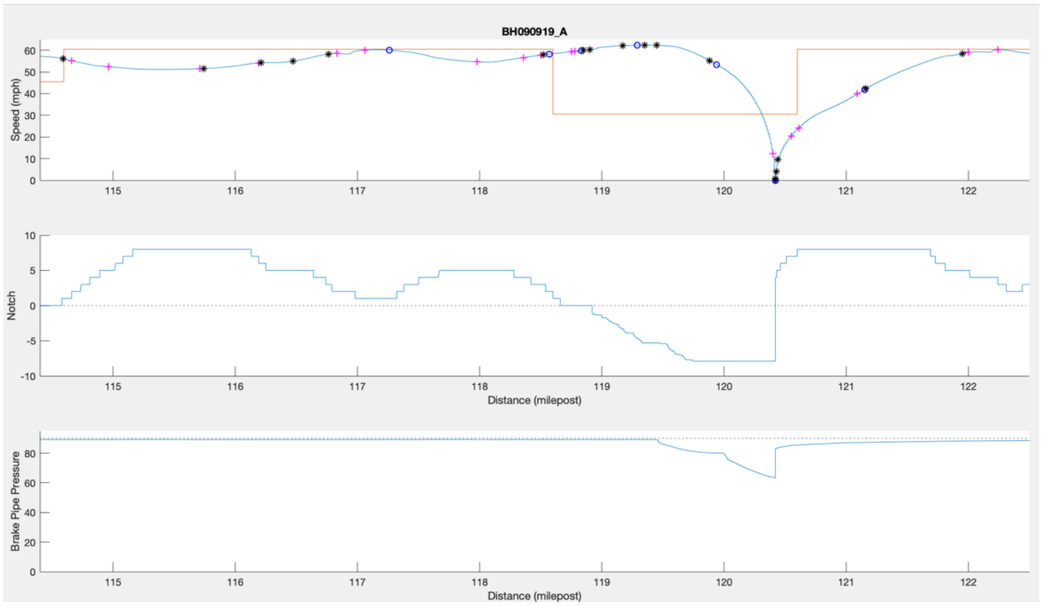


Figure 5-10: Subject pair BH train handling data for Key Decision Point 4 (stop signal). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

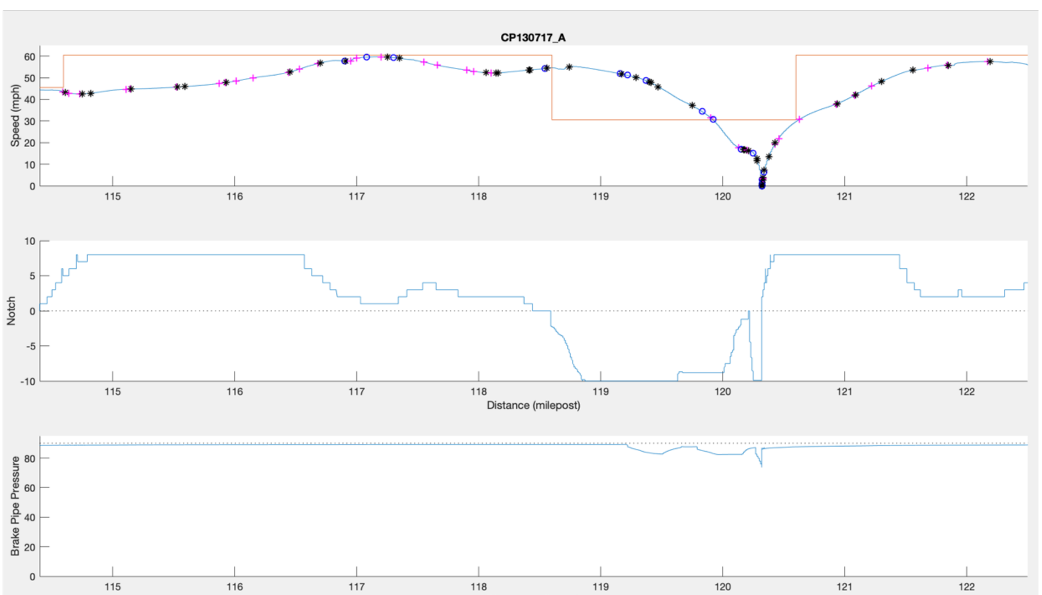


Figure 5-11: Subject pair CP train handling data for Key Decision Point 4 (stop signal). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

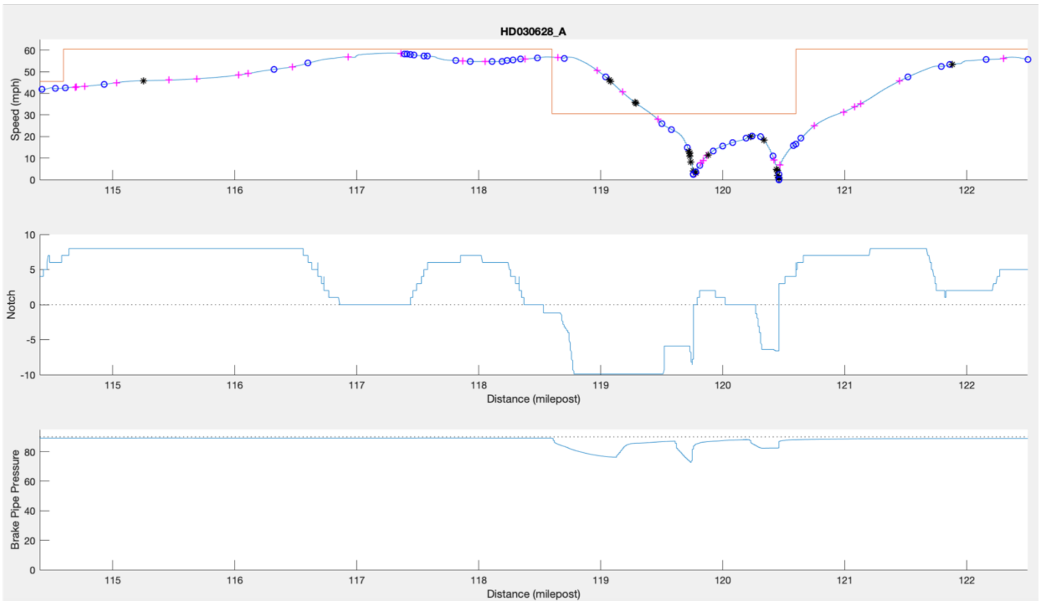


Figure 5-12: Subject pair HD train handling data for Key Decision Point 4 (stop signal). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

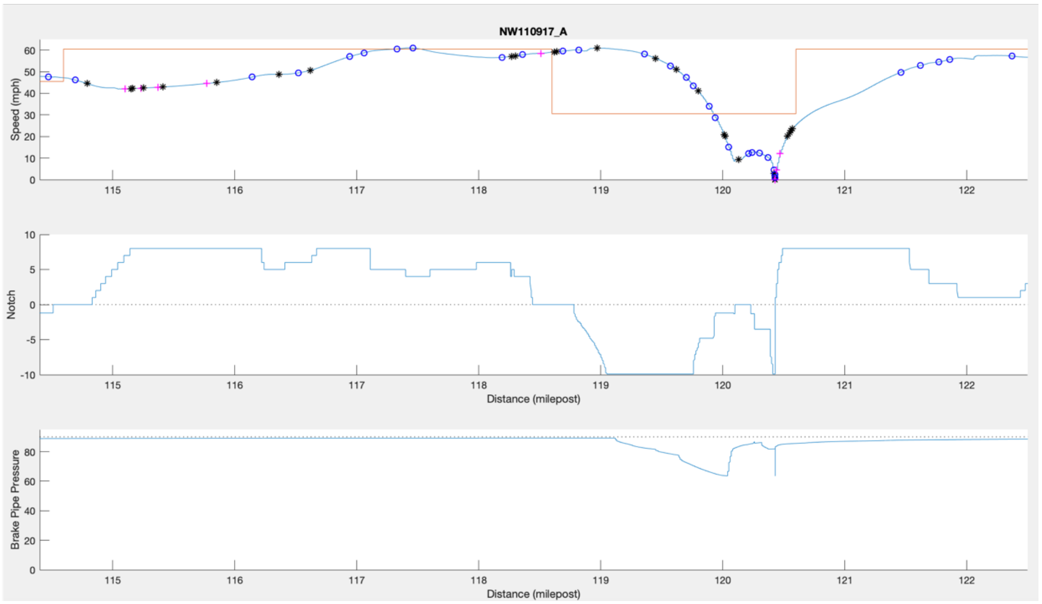


Figure 5-13: Subject pair NW train handling data for Key Decision Point 4 (stop signal). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

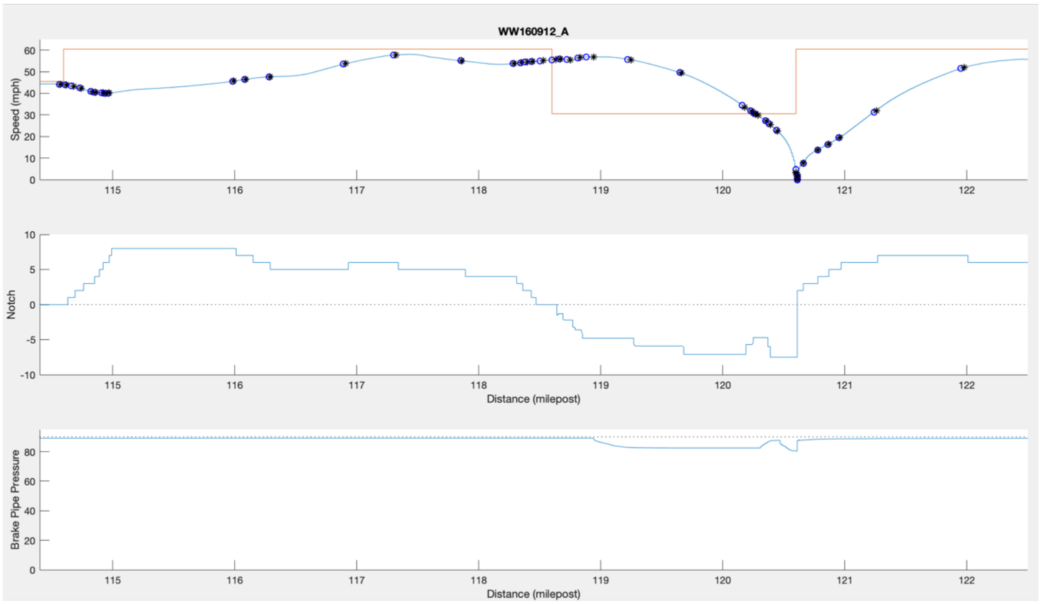


Figure 5-14: Subject pair WW train handling data for Key Decision Point 4 (stop signal). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

Examining the braking usage of each subject pair helps to inform why such variation can be seen. Just after the signal, subject pair HD had both the dynamic and the train brakes applied, which helps to explain HD's early deceleration. Subject pairs CP and WW also applied dynamic brake at or shortly after the signal, but subject pairs BH and NW continued on at idle for a bit past the signal. Train brake usage also varied, with the order of application being roughly HD - WW - NW - CP - BH (HD is earliest application of train brake, BH is latest application of train brake). Note here that subject pair HD applied the air brake almost a mile before BH did.

In addition to location of brake application, it is possible to examine how brakes were applied both quantitatively and qualitatively. Subject pairs BH and NW both had a minimum brake pipe pressure of about 60 psi, which corresponds to heavier brake application and a sharper deceleration. Subject pairs HD and CP had slightly higher minimum brake pipe pressures of about 70 psi. The highest minimum brake pipe pressure was achieved by subject pair WW, who kept the train brake at a relatively constant 80 psi for about a mile and half. Note also that experts varied in the way that the applied the train brakes. Some subject pairs applied the train brake more or less progressively, without coming off of the train brake. This is seen in subject pairs BH and NW, though it should be noted that subject pair NW came to a stop early and had to release and then re-apply train brake. Subject pairs HD and CP both applied train brake and released train brake multiple times while coming to a stop, though again subject pair HD came to a stop too early and was forced to inch the train forward. Interestingly, subject pair WW applied a very small amount of train brake and held that constant for the majority of the 2-mile section between the yellow and red signals.

Clearly, there is much variation in how expert subjects decided to bring the train to a stop, and there are many aspects that could be examined. There is little consistency in strategy between the subjects - just because two subjects adopted similar strategies for one aspect of braking does not mean that they adopted similar strategies for other aspects of braking. This means it is difficult to classify each engineer in terms of an overall braking strategy. However, it is notable that subject pairs HD and CP tended to be well-aligned in their chosen strategies, as did subject pairs NW and BH. Subject pair WW could not consistently be grouped with either of those two sets.

As will be discussed in Chapter 6, there are a few main points stemming from this analysis. First, engineers vary greatly in their desired speed profiles, as evidenced by the variation in where engineers actually began slowing the train. As a result, the location of braking and the nature of the braking also varied. Different engineers also used different combinations of dynamic and train braking to achieve their goals. Furthermore, the lack of consistency between engineers means that it is difficult to identify distinct overall driving strategies from amongst the five subjects.

5.2.2 Key Decision Point 5, NAC Scenario

The second set of plots generated were focused on Key Decision Point 5, which is a Temporary Speed Restriction (TSR) from MP 130 to 131. Like the other TSRs on this route, the speed limit for this TSR was 45 mph, the TSR itself lasted for one mile, and there was a warning flag posted two miles before the TSR, at MP 128.

Based on the results from KDP 4, the first points of comparison for this analysis center on when and how brakes are applied and the final speed profiles for the trains. Notably, subject pair BH failed to recognize the TSR at all (or, in fact, any of the other TSRs along the NAC route), so discussion will focus only on the remaining four subject pairs.

Interestingly, while all subject pairs used both dynamic and train brake to bring the train to a stop for KDP 4, no subject pairs used train brake for slowing the train in KDP 5. In fact, only subject pairs CP and HD even used the dynamic brake, while subject pairs NW and WW never dropped below idle between MP 127.5 and MP 132.5. For subject pairs that did use the dynamic brake, the brakes were not applied at similar points along the route. Subject pair CP applied slight dynamic brake half a mile after passing the warning flag, while subject pair HD applied a heavier dynamic brake closer to the actual speed restriction.

The speed profiles vary noticeably for KDP 4 as well. A useful metric here is the speed at which each subject pair entered the TSR. Both subject pairs CP and HD entered the TSR (MP 130) below the prescribed 45 mph. Subject pair WW was going almost exactly 45 mph at MP 130, and subject pair NW was overspeeding slightly upon entering the TSR. It is also interesting to see that subject pair WW was going much slower at MP 128 (the TSR warning flag) than all of the other subjects, and was able to more or less maintain speed until the TSR zone instead of having to slow down significantly.

The lack of brake application compared to KDP 4 makes it difficult to dissect different aspects of braking strategy. However, the main themes that were present in KDP 4 appear in KPD 5 as well. Engineers vary in when and how to apply braking action, which influenced the final speed profile and whether or not engineers were successful in entering the TSR at the requisite speed.

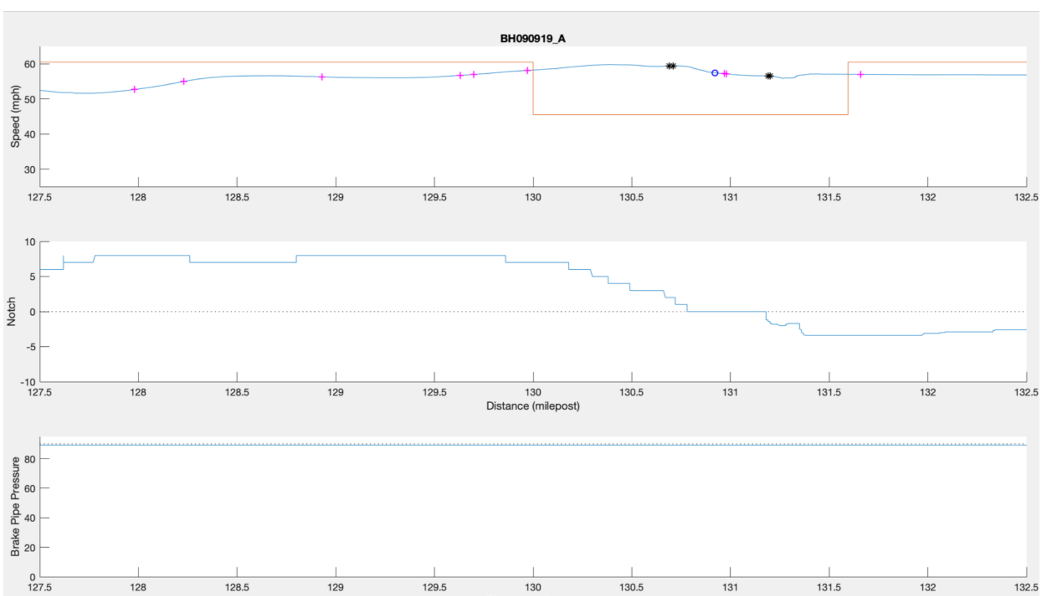


Figure 5-15: Subject pair BH train handling data for Key Decision Point 5 (temporary speed restriction). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

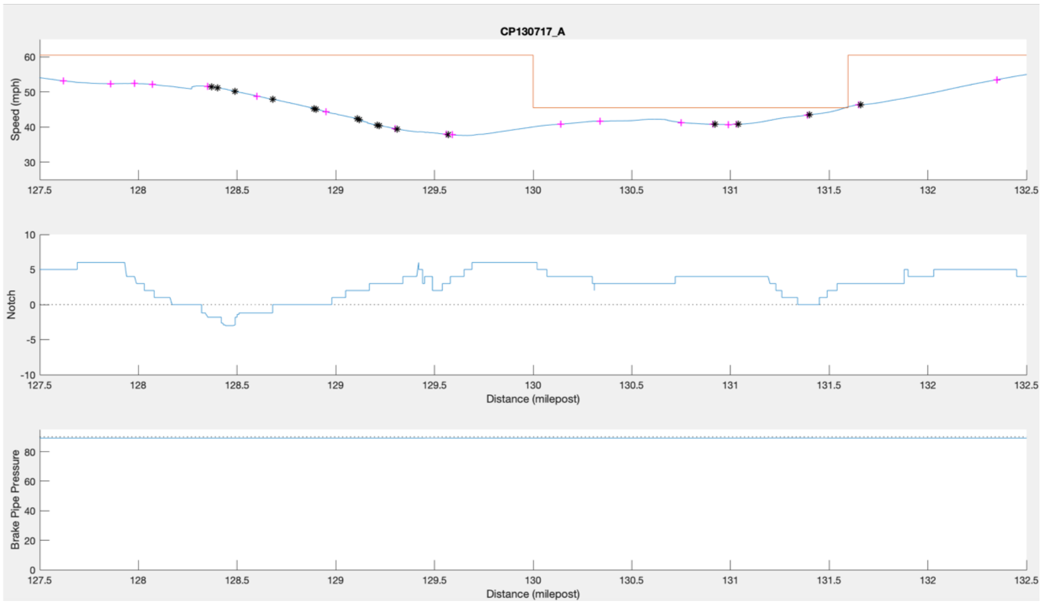


Figure 5-16: Subject pair CP train handling data for Key Decision Point 5 (temporary speed restriction). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

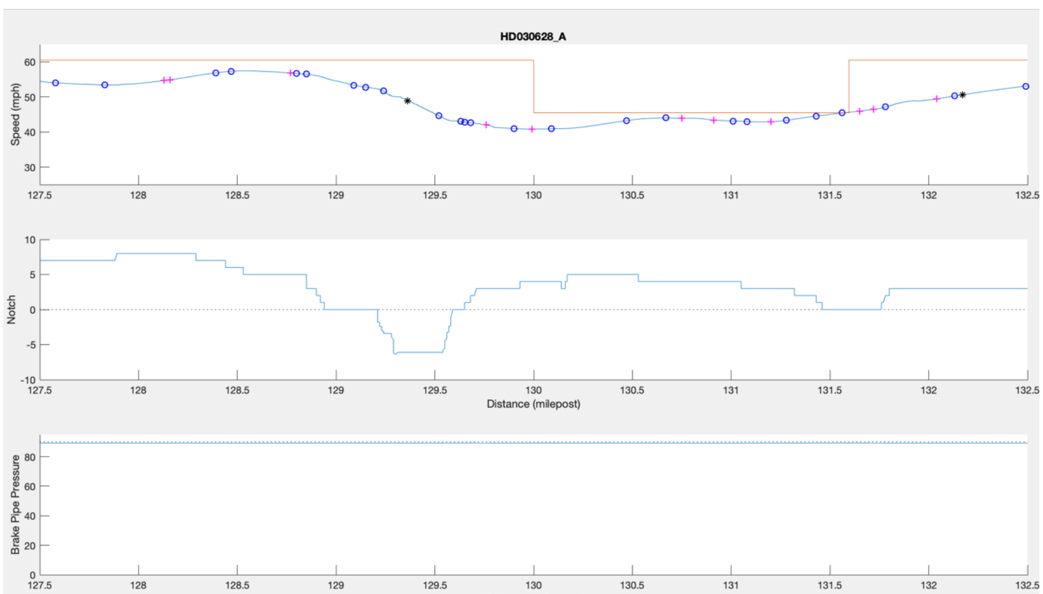


Figure 5-17: Subject pair HD train handling data for Key Decision Point 5 (temporary speed restriction). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

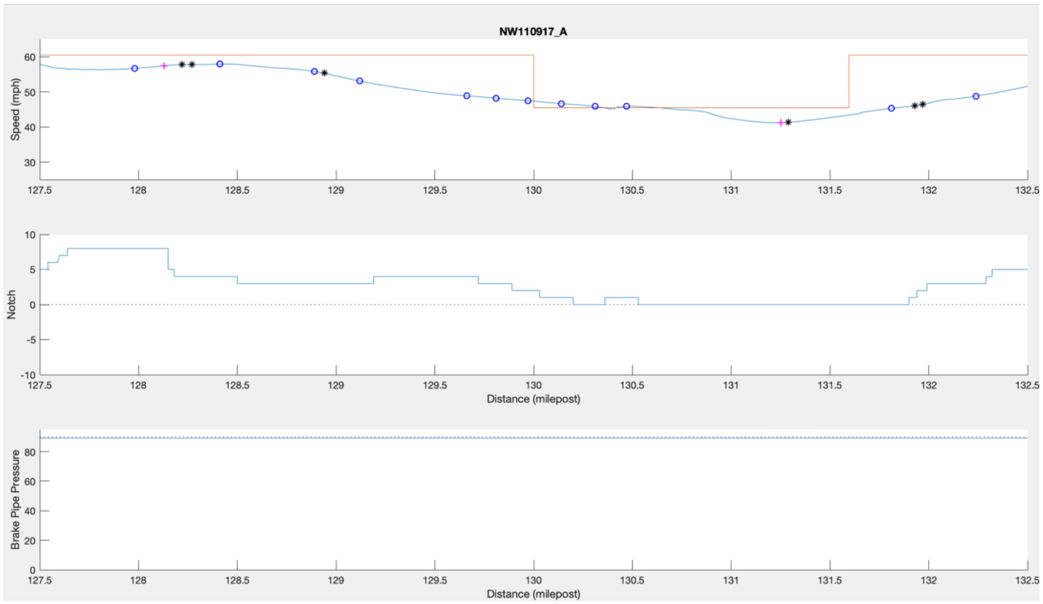


Figure 5-18: Subject pair NW train handling data for Key Decision Point 5 (temporary speed restriction). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

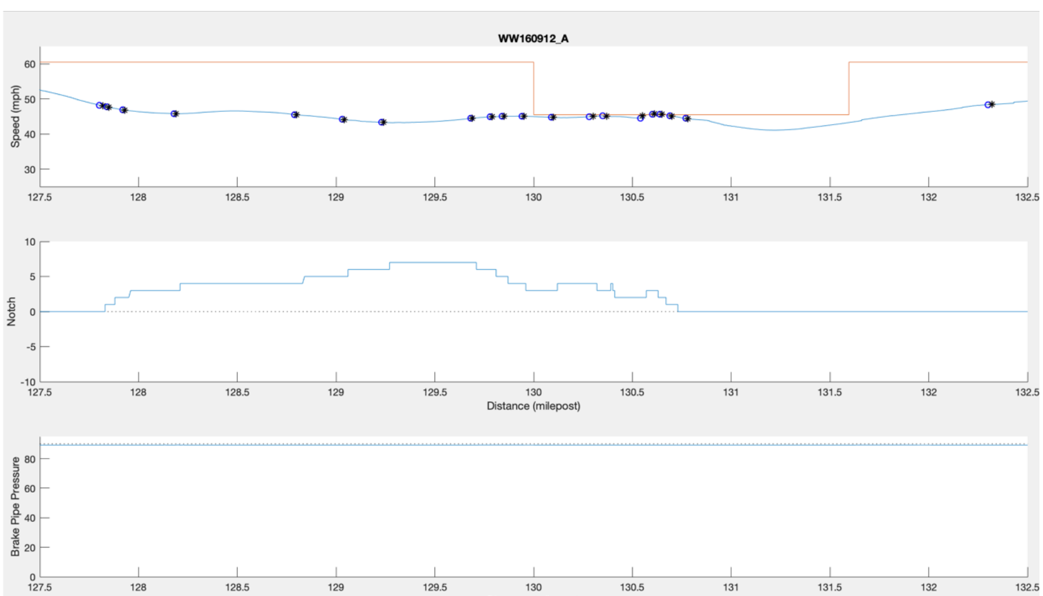


Figure 5-19: Subject pair WW train handling data for Key Decision Point 5 (temporary speed restriction). Top plot is actual speed vs speed limit with interactions marked by various symbols, middle plot is notch/dynamic brake, and bottom brake is brake pipe pressure. A decrease in brake pipe pressure denotes that brakes have been applied, larger decreases indicate higher braking force.

Chapter 6

Implications for Mental Models

The term "mental model" broadly refers to the human's ability to understand and predict behavior of real world systems. In the context of train handling, the engineer's "mental model" incorporates an assessment of the overall situation, and knowledge of train operating rules and behavior in response to control inputs. Adapting terminology proposed by Endsley (Endsley 1995), the engineer monitors the train state and overall situation - as defined by available cues such as the train's location, speed, acceleration, visible trackside signals, and upcoming speed restricting signals or signboards and information radioed by a dispatcher. The engineer seeks to match available cues with a mental model "schema". These are often conceptualized as a set of "if-then" rules defining different situations, what control actions to take to accomplish the goals, and what the result of the actions will be. These schema presumably derive from prior training and experience, the railroad's operating rules and practices and the operator's higher level goals. The following sections describe the elements of the engineer's mental model schema deduced by analysis of experimental train handling data, and discuss the significance of the variability in behavior between the novice/expert teams.

6.1 Decision-Making Inferences and Driving Strategy

The high frequency of the "check precondition" interaction, as detailed in Section 5.1.1, suggests that much of the freight engineer's strategy rests on sets of preconditions that determine the timing and type of actions that are made to accomplish the task goals. The situation is continually re-assessed to determine if additional preconditions have been met and actions must be taken at that time. High-level decisions such as when to begin slowing the train are generally planned in advance, as are the set of preconditions that trigger the action plan. The driving strategy of a particular engineer will be determined by their own internalized goals, but the frequency and consistency of the "check precondition" interaction suggests, as Endsley suggests, that repeatedly checking cues to see if they satisfy preconditioned schema is fundamental to train handling behavior.

6.1.1 Required Mental Model Inputs

As discussed in Chapter 5, it was hoped that "query" interactions during the EAC scenario would reveal what cues expert engineers were looking for and how often they needed them to decide whether the situation matched the requirements for a specific schema. Because the expert engineer could only obtain forward visual cues indirectly by querying the novice conductor, the content of their queries, encoded by subcodes of milepost, signal, grade, and speed restriction, represent the information that was most frequently sought. Mileposts provide spatial information that allowed the engineer to plan their actions for upcoming decision points like signals or speed restrictions. Grade information informed the engineer's expectations with respect to train handling. Signals and speed restrictions had to be observed in order to continue the route safely. Together, it was assumed these verbal query derived cues would help the engineer determine specific actions to drive safely and efficiently.

However, the relatively low frequency of the query interactions seen in the EAC scenarios suggests these codes were not a good proxy for the true frequency at which the engineer sampled their situation. Two factors likely contributed to this. First, during EAC scenarios experts were permitted to give standing reporting instructions to novices, so they would not have to repeatedly ask “what milepost are we at?” or “do you see a signal?”. This way the novice automatically provided basic information e,g,, mileposts, or signal state. Second, post-experiment debriefing with experts revealed that experienced engineers did not rely exclusively on the novice for grade information. Some preferred to infer grade by monitoring the operating display speedometer, which also showed speed gained or lost (mph per minute) based on current acceleration.

Notably, one of the five expert subjects (subject NW) queried the novice for external information much less frequently than the other four. In the post-experiment debriefing, this expert said that while he was aware that he could request external information from the novice, he did not feel it was necessary. However, unlike the others, his EAC scenario data showed multiple significant violations of the speed limits.

6.1.2 Variation Between Engineers

For nominal operations, one of the engineer’s goals is to stay below all applicable speed limitations, including track speed and temporary or permanent speed restrictions. However, these speed limits do not fully define determine the train handling strategy and the resulting train speed trajectory. As illustrated by the comparison plots from Key Decision Points 4 and 5 (Figures 5-10 through 5-19), and discussed in Section 5.2, NAC scenario teams employed different driving strategies when slowing in advance of a stop signal (KDP 4) or a temporary speed restriction (KDP 5). Because the subject pairs tended to combine different aspects of braking strategy in different ways, there is no one useful classification process that can be used. There are many reasons that engineers might choose to employ a specific strategy or sub-strategy. Subject pairs that braked later (or in the case of KDP 5, not at all) might have placed a higher priority on minimization of trip time and were confident in their ability to delay braking, which allowed them to maintain higher average speeds. Subject pairs that braked earlier, which often coincided with a higher frequency of control manipulation, may have assessed their situation more frequently, allowing them to control the train more smoothly by manipulating the controls more often as the situation evolved. In terms of how brakes were applied, use of the dynamic brake might indicate a desire for finer control, since the dynamic brake can be applied and released incrementally, unlike the train brake (which must be fully released to achieve any reduction in braking). Other aspects of train handling might also come into play. For example, engineers may make an initial small application of both dynamic and train brake in order to "bunch" the train when stopping at a signal, in order to reduce run-in forces on train couplers. Finally, engineers that brake early or undershoot the target speed and have to bring the train back up to speed might have less confidence in their mental model and ability to project the situation, and so will attempt to achieve ultimate speed goals more quickly to ensure they do not overspeed.

The differences in preferred combinations may also arise due to external factors. Interviews with some subject matter experts during the experimental design phase suggested that experienced engineers often prefer to avoid using the train brake when possible. However, during the study's post experiment debrief, another engineer commented that he generally would not use dynamic brake and would rely heavily on the train brake in normal operations. This expert typically drove much shorter and lighter trains than the other experts, as he worked for a short line railroad, instead of a Class I railroad. The expert explained that dynamic brakes on his railroad's locomotives were unreliable, and that he believed it was beneficial to train novice engineers to master the train brake rather than permit them to rely on the more easily adjusted dynamic brake. Thus, experienced engineers vary not only in their projected/preferred speed profiles but also the tactical decisions required to achieve those profiles.

6.2 Implications for Mode Design

As discussed in earlier sections, the primary purpose for evaluating train handling performance and verbalizations that might reveal engineer mental model schema was to inform design of an enhanced automation system in freight locomotive cabs. For safer and more efficient operation of automated modes, engineers should be able to understand and predict the behavior of the automated system. If the automation behavior deviates substantially from the engineer's expectations, the engineer has to dedicate mental effort toward understanding the reason for the automation's behavior, distracting them from their primary responsibilities. Alternatively, the engineer might simply avoid using the automation. The latter behavior was seen with initial introduction of the GE Trip Optimizer autothrottle system into locomotive cabs. Freight engineers sometimes commented that though they knew the Trip Optimizer speed profile saved fuel, the profile was different than the one they would have used themselves, and they did not always understand why. For example, TO might coast in an idle throttle setting one or two miles in advance of a speed restriction in order to bring speed down without braking, whereas the human engineer would ordinarily continue at a higher throttle setting for longer time before slowing down much closer to the speed restricted area. Most engineers habitually try to maintain a high average speed in order to stay somewhat ahead of schedule, as a hedge against later contingencies. Initially, engineers frequently disabled TO when the automation seemed to be acting in a way the engineer perceived as undesirable [1].

6.2.1 Automation Teaming

The implication of the data from this study is that engineers largely act as situation assessors. In other words, train driving skill requires the execution of a pre-defined set of actions pursuant to a set of preconditions, and it is the engineer's job to identify these preconditions and determine if they have been met. Encoding and remembering

[1] Personal communication with James Brooks, research engineer at GE who worked closely with companies integrating TO into their locomotives.

a set of “if-then” clauses is a fairly straightforward task for both machines and people, but evaluating cues and determining the truth of the “if” statement is significantly more difficult, requiring integration of multiple stimuli. This implies that a human operator’s role should center on the synthesis of external and internal information to assess a situation and select an appropriate mode that triggers the expected action.

6.2.2 Engineer Inputs

In the design of a new type of automation that is consistent with an engineer’s normal driving strategy, one must ask: What parameters in the automation strategy should an engineer be able to adjust? When should the automation prompt the engineer for more information?

Since there is reason to believe that engineers have individual internal goals that result in different observed train control behavior, they should have some ability to influence the automation’s planned speed profile. As discussed previously, there are many aspects of braking strategy, and engineers that choose a similar approach for one aspect of braking strategy may choose different approaches for other aspects. This implies that engineer inputs may need to be multi-dimensional, not simply "drive more aggressively" or "drive more efficiently", since driving strategies vary along more than one axis. One potential solution would be to simply allow the engineer to directly adjust the automation control laws and have the automated system issue a warning if the engineer’s desired inputs would be impossible to safely achieve. For instance, if the automation speed profile indicates that it will begin to brake immediately after passing a TSR warning flag, the engineer might be able to select an “idle” option to let the train coast until closer to the start of the actual speed restriction, thereby shaping the final braking profile. If the engineer selects a less efficient speed profile, e.g., a profile that uses more fuel, the automation might warn the engineer that the desired changes would be less efficient, but still permit the changes. On the other hand, if the engineer is attempting to input an unsafe speed profile, then the automation might inform the engineer of the invalid speed profile and default to the closest speed profile that permits safe operation of the train. Some consideration would still need to be given for extraordinary circumstances where conditions might require a speed profile that would otherwise be considered unacceptable.

6.3 Experimental Limitations

While the verbal interaction coding implementation was useful for high-level qualitative analysis, in its current form, it is difficult to draw specific conclusions from the coding that informed design specifications for the enhance control mode, other than that to accommodate differences in goals and driving style between engineers, the modes should be designed so that they have user adjustable parameters – such as braking aggressiveness – whose effects are easily predicted.

One of the primary limitations of the coding method was that the engineers did not verbalize all of the information gathering or processing that they actually performed - for instance, there were relatively few Level 2 “evaluate’ interactions in this category, so the measured frequency of evaluation events may actually significantly underestimate the true frequency. The engineers were probably drawing information from other sources such as the paperwork and train displays and the visual sampling from these sources was not monitored. Also, since the study wound up testing only five subjects, and analysis focused on only two decision types (responses to signals and slow orders) many other situational train handling strategies doubtless remain to be identified. The highest number of notch and brake adjustments necessarily occurred as the train approached areas requiring speed changes. Though the behavior could be characterized somewhat qualitatively, it was not possible to directly assess the reasons for the behavior, which was presumably associated with higher level strategic goals of the engineer.

There were also some limitations inherent to the data visualizations used that were based on distance (in milepost) and not time. The areas of very high density activity in the whole-route histograms are where the train was slowing to a stop. As a result, interactions that would be evenly spaced in time are not evenly spaced in distance, since a slower-moving train covers less distance in the same time interval.

There are a number of limitations inherent to a simulation-based study. Physical cues such as acceleration cues, inclination of the track, and the movement of the train cars (e.g., bunching and stretching) are not reproduced in the CTIL simulator and these cues likely affect the handling of the train. There was also no real threat to safety nor any risk associated with more aggressive handling of the train, so engineers might behave differently than they would if they were driving along the mainline. Though subjects were instructed to follow normal operating rules and procedures as closely as possible, sometimes they chose not to. For instance, usually a short "set-up" period is required between selecting a throttle setting and selecting a dynamic brake setting (and vice versa) in order to protect the engine. While instructing the novice, one expert subject explained that usually this period would be required, but since the cab was simulated and there was no real locomotive engine to damage, there was no need to include the set-up period. On the other hand, the fact that subjects were aware that their actions were being recorded and analyzed could also have led some to deviate from their normal driving style. As noted in Section 6.1.3, one expert subject differed from conventional wisdom regarding the train brake, preferring it over the dynamic brake. However, this subject also noted he had tried to use the dynamic brake more than he usually would, since he was aware that he relied on the airbrake more than would be typical for a Class I railroad engineer and wanted to emulate a more typical driving strategy. Of course, subjects may have also felt pressure to drive more conservatively, given that any rules violation was likely to be noticed.

Another factor that may have impacted train handling behavior was that the route and the train were unfamiliar to the expert subjects. Generally, a locomotive engineer is extensively trained for a particular route before he or she can drive it on their own. To qualify on a route, an engineer is often asked to draw the route from memory, including the physical characteristics of the environment, signals, and relevant track equipment. Engineers remember route specific "braking points", and associated landmarks to look for. For this experiment, none of the engineers were familiar with the route prior to the expert training run. Even with access to the track chart at all times, they could not achieve a realistic level of familiarity with the route used in the experiment, and this likely influenced their driving behavior to some degree. It is noteworthy that the industry is becoming more interested in "route interoperability" and the possibility of assigning crews to drive routes on which they have significantly less experience. CTIL studies examining the effect of prior route experience on train handling behavior may be particularly helpful in the future.

6.4 Recommendations for Future Experiments

Several changes should be implemented in a future experiment to overcome the limitations outlined above. To elicit a more complete record of all of the expert subject's conscious and unconscious thoughts, test subjects could be instructed to verbalize their thoughts continuously throughout the experiment, including not only physical actions and tactics, but higher level goals and strategies. This method could potentially result in higher frequencies of the interaction types and thought processes that were hypothesized for this experiment. In the present experiment, these verbalizations – such as the "evaluate" interaction - occurred relatively infrequently, Unfortunately there is some likelihood that the subconscious thought processes and considerations that form an important component of the engineer's mental model might not always be verbalized by the subject. Nisbett noted that people struggle to report cues for higher-order processes, and that the accuracy of reported cues tends to be low (Nisbett 1977). For instance, an engineer might be able to say when it is time to start slowing down for a signal, but would have difficulty explaining what exactly made him or her know that was the correct location. The human brain is capable of recognizing situations without using rules. Even if he or she did give the reasoning behind the decision, it is possible that the subject did not correctly identify the mental processes underlying their decision-making process and instead gave an answer based on their own hypothesis on how such decisions should be made. Additionally, there is the concern that the act of verbalizing thoughts might alter an engineer's behavior or distract him or her from the primary task.

An alternate method might require a supervising expert subject to evaluate the performance of an engineer driving a route normally, with no special requirements levied on the driving subject either during or after the driving portion of the experiment. Assuming that an observing expert subject would be familiar with most of the goals and strategies that the driving engineer could be using, the observer could explain the behavior of the driving engineer. The main benefit would be the ability to collect results similar to a single engineer voicing her or her thoughts, but without the added distraction of another task threatening the integrity of the data. However, since mental models vary between engineers, one engineer's explanation of another engineer's train handling may not be entirely accurate. Alternatively, an engineer could report on his or her own actions by watching a video record of the trip, but the subjects might have more difficulty explaining areas where mistakes were made or attempt to retroactively justify a poor decision.

A third potential method to force an engineer at the controls to report or request more information would be to obscure additional critical information, such as the speed or acceleration of the train. In this case, the engineer must actively request the hidden information by pressing a button or lifting a cover to see the desired information. The act of revealing the relevant states might also alter the sampling rate, but this method might give more insight to the inputs for an engineer's mental model from inside the train. If a large number of cues were hidden and several were needed for a particular task, the additional workload to manage the information acquisition conceivably might overload the engineer and degrade primary train control task performance.

The experiment as designed in this thesis did not include any spontaneous or special challenges in terms of train handling strategy. As a result, nothing could be learned about how an engineer assesses a novel situation and adopts their mental model to the new situation. As demonstrated by the dominant "check precondition" interaction, especially compared to interactions that modified the plan in one way or another, the expert subject was largely able to construct all of the necessary actions and preconditions before encountering a situation. Testing an experienced engineer's skills on a more challenging route with steeper grades or an artificial challenge like following a simulated train at a safe distance might also reveal more information about the mental model. The terrain in the track subdivision available in the CTIL does not contain any sections that are difficult to drive. However, the engineer could be given conflicting goals, like a speed restriction awkwardly placed on an uphill grade that would make it difficult for the engineer to make it up the hill without overspeeding. This type of situation would force the engineer to develop a new set of pre-conditions and actions to handle the new task and might shed some light into the decision-making process of an experienced engineer.

Finally, it would also be interesting to see how well an engineer learns to predict the performance of his or her train on an unfamiliar route. An engineer might be given a segment of a track chart for an unfamiliar route before driving the route and asked to annotate roughly where he or she would apply notch or brake, and explain his or her reasoning. This line of questioning would get at the underlying mental model the engineer uses to predict the movement of the train, without the advantage of seeing how the train reacts in real time. Afterwards, the engineer would drive this specific segment. To ensure compatibility between the two trials, both segments would have to start from a known state, like stopped at a signal just before the experimental segment. This method would permit the assessment of how well an engineer can predict performance of a train, thus evaluating the accuracy of the engineer's mental model, as well as create a slow-paced environment for the engineer to walk through his or her mental model. A potential complication would be if the engineer remembered where he or she had annotated certain control inputs and then acted according to those predicted inputs instead of driving normally. Ultimately, no one method will fully elucidate an engineer's mental model – to gain a more complete understanding, multiple methods probably need to be used in conjunction with one another.

Chapter 7

Conclusion

The goal of this study was to enhance the understanding of an expert engineer's mental model of how to drive a freight train, in order to better inform the design of a new supervisory control mode for locomotive operations. To accomplish this, expert locomotive engineers were paired with novice subjects to complete a simulated trip while the interactions between expert and novice were recorded and subsequently studied. In one scenario, the expert subject directed the novice subject, who was manipulating the train controls but was unable to see forward outside of the simulated cab. The expert engineer's directions and explanations were analyzed to understand the context of the control actions made by the novice. In the second scenario, the expert subject manipulated the controls, but had to rely on the novice subject's ability to see outside of the simulated cab. The timing and content of information requests from the expert engineer helped illuminate the specific information they used in their train handling decision-making processes.

A coding scheme was developed to characterize the interactions between the expert and novice subjects, which was particularly relevant for analyzing the NAC (novice-at-the-controls) scenario, where the expert was verbally instructing the novice how to drive the train. The most common interactions between expert and novice for the NAC scenario were classified as "execute" and "check precondition" interactions. The "execute" interaction frequency was unsurprising because it indicated a directive to perform an action, and in this NAC scenario, the expert needed to give directions frequently to the novice. Particularly interesting was the frequency of the "check precondition" interaction. Although recorded verbal interactions may not perfectly mirror an expert's thought process, the high frequency of "check precondition" interactions suggest much attention is devoted to evaluating situational cues and determining if a specific situation meets those preconditions.

Generally, it appears that an engineer determines what decisions and actions should stem from meeting different preconditions, then monitors the situation to determine which preconditions are met. As a result, the engineer's primary role isn't solely to make high-level decisions. Instead, the engineer's usefulness is his or her ability to assess a situation accurately, determine what pre-conditions have been met, and execute the decision that is associated with the set of pre-conditions. For the proposed supervisory control mode, this is an important distinction that informs the overall system design. The automation can be pre-programmed to make decisions in response to a particular situation, but it needs the operator to correctly assess a situation and inform the automation. The system should also prompt an operator for input when the system lacks the information to determine whether pre-conditions have been met.

In looking at the interaction frequencies and types along with train handling data for different operators in the NAC scenario, large variations in both interaction style and driving style as instructed by the expert were clearly evident. For instance, when approaching a red signal, the subject teams varied on how and when to apply braking action. There are a few potential explanations for these differences. It is possible that some engineers are more confident in their mental model and ability to handle the train, and thus are more comfortable maintaining a higher speed for longer. Such engineers might also be attempting to save some time by maintaining a higher speed. Others engineers might prefer a more active management of the train, manipulating the controls earlier and more frequently. Different engineers used the dynamic and train brakes differently as well, which might be influenced by train handling characteristics like bunching and stretching. Similar differences in strategy were also seen in the interaction versus train handling plots for a speed restriction scenario, with some engineers preferring to immediately reduce speed at the warning flag, while others preferring to attain the restricted speed right as the train entered the speed restricted area. Because each engineer combined different aspects of braking strategy in different ways (e.g. not all engineers that braked early used the same amount of train brake, as evidenced by minimum brake pipe pressure), it is difficult to classify braking strategy into broad groups.

These differences mean that even if two engineers share a strategic, high-level goal (for example, stop before the red signal), there will be differences in each engineer's tactical decisions to achieve that goal through manipulation of the controls. The differences in their tactical decisions are reflected in the resulting braking curves and use/non-use of the air brake. For the supervisory control mode, this suggests that engineers may prefer to influence the tactical decisions made by the system, in addition to setting high-level goals through situation assessment. Thus, the supervisory control mode should allow some fine-tuning, so that an engineer can more closely match the system's behavior to his or her own mental model. The experiment described in this work suggests that braking profiles are one parameter that the engineer might wish to control, though there are other aspects of control strategy that this work did not analyze.

The analysis done for the EAC (expert at the controls) scenario was intended to reveal the types of information an expert engineer relies on to build their mental model, and how frequently an expert updates that mental model. Specifically, query interaction codes identified the engineer when an engineer asked for more information. This would then indicate the information needed at that point in the scenario and how often the engineers felt they needed more information. Unfortunately, the study engineers used these interaction types so infrequently, that it was not possible to confidently assess the quantities of interest.

Interestingly, the lack of requests for information does reveal some information about other sources of information the engineer uses to assess a situation. During the experimental scenarios, the expert subject was equipped with a bulletin listing the mileposts beginning and ending speed restrictions, and track chart that showed crossing and signal locations but was stripped of grade information. The experts were permitted to give standing orders for the novice observer to report mileposts, signals, and signage when they appeared. This set of information permitted the expert subjects to drive a large portion of the route safely, supplemented by queries to the novice regarding the current state of the train. However, one expert subject that refrained from asking the novice for any additional information was also the most prone to overspeeding, suggesting that the additional expert-to-novice queries, despite their infrequent use, were important to safely operate the train.

Looking at codes which identified the subject of queries, the most common were for milepost and signal, with grade and speed restriction also prominent query types. While these results are not unexpected, they confirm the current understanding of the types of information required to accurately assess a situation.

This interaction coding is a new method for identifying the some of the engineer's mental model constructs, and to the best of the research team's knowledge, has not been previously attempted in rail research. The primary limitation of this method is that current research suggests experts often recognize situations using subconscious thought processes, and know what actions to take, and are sometimes unable to explain why without retrospection. Hence one limitation of verbalization analysis is that some factors that implicitly determined the expert's actions are not apparent in the contemporary verbalizations. Another approach might have the expert subject verbalize everything that comes to mind, which could lead to a more complete picture about how an engineer's situation assessment evolves over time naturally. Other alternative methods might include a retroactive commentary by an engineer after driving a route, or an observer providing commentary while watching another engineer drive a route.

Overall, the experiment yielded some empirical information about the contents and processes of an expert freight engineer's mental model that can be incorporated into the development of a more flexible supervisory control mode. Improvements to the methods used here will hopefully provide better quantitative information about these mental models that inform the design and operation of new automation. This will be needed as locomotive cab automation becomes more capable of responding to changes in the operating environment.

Chapter 8

References

Association of American Railroads. (2020). "Positive Train Control." Accessed May 7, 2020. https://www.aar.org/wp-content/uploads/2020/01/AAR-PTC-Fact-Sheet-2020.pdf

Bainbridge, Lisette. (1983). “Ironies of Automation.” *Automatica* 19, no. 6: 775–779.

Branton, Paul. (1978). "The Train Driver." In *The Analysis of Practical Skills*, 169-188. Springer, Dordrecht.

Brooks, James, et al. (2016). “Design and Evaluation of a Robust Manual Locomotive Operating Mode,”, unpublished research project proposal, Volume I. GE Global Research and MIT Man Vehicle Laboratory, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. Abstract: https://cms8.fra.dot.gov/human-factors/ctil/browse-research

Chermack, Thomas J. (2003). Mental Models in Decision Making and Implications for Human Resource Development. *Advances in Developing Human Resources* 5, no. 4: 408-422.

Coffey, John W., Robert Hoffman, and Alberto Cañas. (2006). "Concept Map-Based

Knowledge Modeling: Perspectives from Information and Knowledge Visualization." *Information Visualization* 5, no. 3: 192-201.

Core, Mark G., and James Allen. (1997). "Coding Dialogs with the DAMSL Annotation Scheme." In *AAAI Fall Symposium on Communicative Action in Humans and Machines* 56, pp. 28-35.

Clark, Robert E., and Fred Estes. (1996). "Cognitive Task Analysis for Training". *International Journal of Educational Research* 25, no. 5: 403-417.

Eldredge, David, and Paul Houpt. (2014). "Trip Optimizer for Railroads." Accessed May 7, 2020. http://ieeecss.org/sites/ieeecss/files/2019-06/IoCT2-SS-Eldredge-1.pdf

Elias, Bart. (2019). "Cockpit Automation, Flight Systems Complexity, and Aircraft Certification: Background and Issues for Congress". Report No. R45939. Congressional Research Service.

Endsley, Mica R. (1995). "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors* 37, no. 1: 32-64.

Endsley, Mica R., and Esin O. Kiris. (1995). "The Out-of-the-Loop Performance Problem and Level of Control in Automation." *Human Factors 37*, no. 2: 381-394.

Federal Railroad Administration. (2011). Railroad Industry Modal Profile. Report No. DOT/FRA/ORD-11/20. Washington, DC: U.S. Department of Transportation, Federal Railroad Administration.

Groshong, Hannah. (2016). "Task Modeling and Assessment for Human-Sytem Interaction in Freight Rail Operations." Master's Thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics. http://library.mit.edu/item/002530075

Houpt, Paul K., et al. (2009). "Optimal Control of Heavy-Haul Freight Trains to Save Fuel." *9th International Heavy Haul Association Conference.*

McGonigal, Robert S. (2006). “Dynamic Braking: Power for the Downgrade.” Accessed May 7, 2020. https://trn.trains.com/railroads/abcs-of-railroading/2006/05/dynamic-braking

Naweed, Anjum, Ganesh Balakrishnan, and Jillian Dorrian. (2018). "Going Solo: Hierarchical Task Analysis of the Second Driver in “Two-Up”(Multi-Person) Freight Rail Operations." *Applied Ergonomics* 70: 202-231.

Nikolic, Mark. I., Nadine Sarter. (2007). “Flight Deck Disturbance Management: a Simulator Study of Diagnosis and Recovery from Breakdowns in Pilot-Automation Coordination”. *Human Factors* 49, no. 4: 553-563.

Nisbett, Richard E., and Timothy D. Wilson. (1977). "Telling More than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84, no. 3 (1977): 231.

Proctor, Robert W., and Trisha Van Zandt. (2008). Human Factors in Simple and Complex Systems, 2nd ed. Boca Raton, Florida: CRC Press.

Roth, Emilie, Hadar Rosenhand, and Jordan Multer. (2013). Using Cognitive Task Analysis to Inform Issues in Human Systems Integration in Railroad Operations. Report No. DOT/FRA/ORD-13/31. Washington, DC: U.S. Department of Transportation, Federal Railroad Administration.

Sheridan, Thomas B. (2011). “Adaptive Automation, Level of Automation, Allocation Authority, Supervisory Control, and Adaptive Control: Distinctions and Modes of Adaptation". *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 41, no. 4: 662-667.

United States Government Accountability Office. (2019). Freight Trains are Getting Longer, and Additional Information is needed to Assess Their Impace. Report No. GAO-19-443. Washington, DC: U.S. Government Accountability Office.

Appendix A

Route Descriptions

The following two pages are the route descriptions for the experiment conducted in the CTIL. The route is part of the BNSF Aurora subdivision driven in an east-to-west direction.

Route Descriptions

Expert Training:
1) MP 77.7 Before grade crossing (signal only)
2) MP 77.9 Turnout to MAIN 2, Signal shows DIVERGING CLEAR aspect
3) MP 82.2 Permanent speed restriction: 40 MPH
4) MP 83.7 Permanent speed restriction: 35 MPH
5) MP 83.9 Permanent speed restriction: 45 MPH
6) MP 84.3 Signal shows APPROACH aspect
7) MP 84.4 End permanent speed restriction
8) MP 86.15 Signal shows DIVERGING CLEAR aspect
9) MP 86.3 Crossover to MAIN 1
10) MP 93.3 Signal shows APPROACH aspect
11) MP 95.8 Begin permanent speed restriction: 45 MPH
12) MP 95.9 Signal shows STOP aspect
13) MP 102.3 End permanent speed restriction
14) MP 138.6 Sign shows APPROACH MEDIUM aspect
15) MP 142.0 Begin permanent speed restriction: 35 mph
16) MP 142.4 Signal shows APPROACH aspect
17) MP 143.2 Signal shows STOP aspect (end route)

Novice-at-the-Controls Scenario (NAC)
1) MP 77.7 Before grade crossing (signal only)
2) MP 77.9 Turnout to MAIN 2, Signal shows DIVERGING CLEAR aspect
3) MP 82.2 Permanent speed restriction: 40 MPH
4) MP 83.7 Permanent speed restriction: 35 MPH
5) MP 83.9 Permanent speed restriction: 45 MPH
6) MP 84.3 Signal shows APPROACH aspect
7) MP 84.4 End permanent speed restriction
8) MP 86.15 Signal shows DIVERGING CLEAR aspect
9) MP 86.3 Crossover to MAIN 1
10) MP 87.0 Begin TSR: 45 MPH
11) MP 88.0 End TSR
12) MP 93.3 Signal shows APPROACH aspect
13) MP 95.8 Begin permanent speed restriction: 45 MPH
14) MP 95.9 Signal shows STOP aspect
15) MP 102.3 End permanent speed restriction
16) MP 105.0 Dispatcher call for meet and pass
17) MP 113.0 Begin TSR: 45 MPH
18) MP 114.0 End TSR
19) MP 115.0 Meet and pass oncoming train
20) MP 118.6 Signal show APPROACH aspect
21) MP 120.6 Signal shows STOP aspect
22) MP 130.0 Begin TSR: 45 MPH

23) MP 131.0 End TSR
24) MP 138.6 Sign shows APPROACH MEDIUM aspect
25) MP 142.0 Begin permanent speed restriction: 35 mph
26) MP 142.4 Signal shows APPROACH aspect
27) MP 143.2 Signal shows STOP aspect (end route)

Engineer-at-the-Controls Scenario (EAC)
1) MP 77.7 Before grade crossing (signal only)
2) MP 77.9 Turnout to MAIN 2, Signal shows DIVERGING CLEAR aspect
3) MP 80.0 Begin TSR: 45 MPH
4) MP 81.0 End TSR
5) MP 81.9 Dispatcher call for meet and pass
6) MP 82.2 Permanent speed restriction: 40 MPH
7) MP 83.7 Permanent speed restriction: 35 MPH
8) MP 83.9 Permanent speed restriction: 45 MPH
9) MP 84.3 Signal shows APPROACH aspect
10) MP 84.4 End permanent speed restriction
11) MP 86.15 Signal shows DIVERGING CLEAR aspect
12) MP 86.3 Crossover to MAIN 1
13) MP 91.9 Meet and pass oncoming train
14) MP 95.8 Begin permanent speed restriction: 45 MPH
15) MP 101.0 Signal shows APPROACH aspect
16) MP 102.3 End permanent speed restriction
17) MP 104.0 Signals shows STOP aspect
18) MP 118.0 Begin TSR: 45 MPH
19) MP 119.0 End TSR
20) MP 122.4 Signal shows APPROACH aspect
21) MP 123.9 Signal shows STOP aspect
22) MP 132.0 Begin TSR: 45 MPH
23) MP 133.0 End TSR
24) MP 138.6 Signal shows APPROACH MEDIUM aspect
25) MP 142.0 Begin permanent speed restriction: 35 MPH
26) MP 142.4 Signal shows APPROACH aspect
27) MP 143.2 Signal shows STOP aspect

Appendix B

Communication Guidelines

The following two pages are the communication guidelines for the experiment conducted in the CTIL. These were reviewed with the subjects prior to the start of the first experiment session.

Communication guidelines: Novice-at-the-Controls

***The expert subject will conduct all communications with dispatcher during this trial** (no need to press any buttons to call dispatch, the mic is always on so speaking loudly is sufficient)

Expert subjects may:

- Give commands with brief explanation
  - Examples: increase power, we are below track speed; decrease speed, we are approaching a temporary speed restriction
- Give standing orders with brief explanation
  - Example: maintain speed until directed otherwise, this is a good pace for us

Briefly and specifically explain the reason for a command or order

  - Good example: We are overspeeding, decrease speed by 3 mph
  - Bad example: We are approaching a maintenance of way and there might be workers on the track, so we will need to call the dispatcher soon and you may need to slow down

Expert subjects are responsible for the quality of the ride and should act as though they are controlling the train through the novice - if the train is not being handled the way you would handle it, give feedback to the novice *immediately.*

Novice subjects may:

- Ask about specific commands that are being executed now or imminently
  - Example: what brake should I be using right now? What speed do you want me to increase to?

Novice subjects should refrain from asking overly general questions, such as “when do you usually start slowing down for a speed restriction that’s coming up?”

Communication guidelines: Engineer-at-the-Controls

***The expert subject will conduct all communications with dispatcher during this trial** (no need to press any buttons to call dispatch, the mic is always on so speaking loudly is sufficient)

Expert subjects may:

- Ask any specific questions about the environment (either outside the train or visible on the TO rolling map)
  - Example: are we currently on a downhill?
- Give specific standing orders
  - Example: let me know anytime you see signage for a speed restriction

Expert subjects are responsible for the overall quality of the ride, and should ask the novice for more information any time that the expert feels they do not have enough information to safely complete the trip. Expert subjects should not ask the novice to infer train performance, however ("are we going to stop before the signal or should I add more brake?")

Novice subjects may:

- Proactively report *only* signals, mileposts, and signage, as well as obstructions on track
- Answer any question from the expert to the best of their ability

Novice subjects should not infer train performance or give strategic handling information proactively. For example, "you will be on an uphill soon, add power to maintain speed", would not be an appropriate statement for the novice during this trial.

Appendix C

Trip Rules

The following page is the standard set of trip rules given to the expert subject for the experiment conducted in the CTIL.

GCOR RULES IN EFFECT ON BNSF Aurora – Steward to Savanna.

BNSF Aurora SUB, Steward to Savanna, IS CONTROLLED BY THE Aurora SUB TRAIN DISPATCHER.

MAXIMUM AUTHORIZED SPEED IS 60 MPH FOR Freight TRAINS ON THE Aurora SUB – Steward to Savanna, UNLESS NOTED BELOW.

PERMANENT SPEED RESTRICTIONS:

MP 82.2 TO MP 83.7	40 MPH
MP 83.7 TO MP 83.9	35 MPH
MP 83.9 TO MP 84.4	45 MPH
MP 95.8 TO MP 102.3	45 MPH
MP 142.0 TO MP 144.5	35 MPH

Maximum Authorized Speed through all crossover turnouts and Siding turnouts is 35 mph

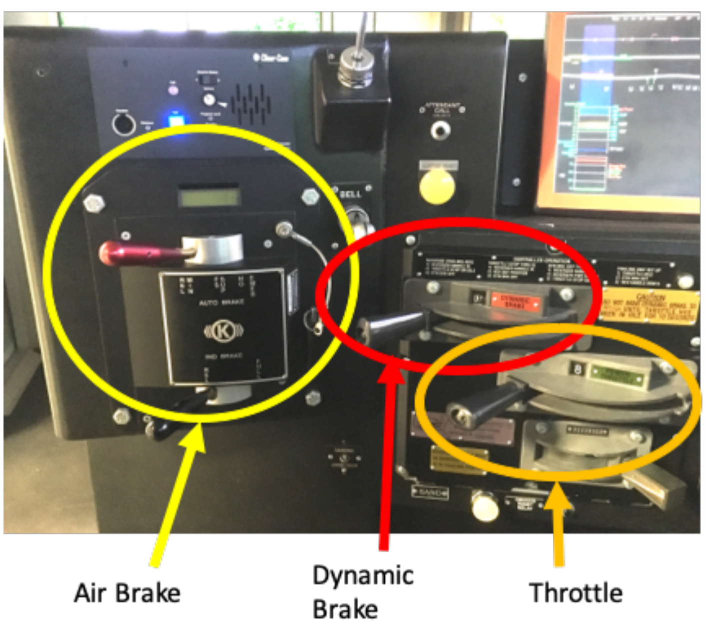
QUIET ZONES FROM MP 77 to MP 145

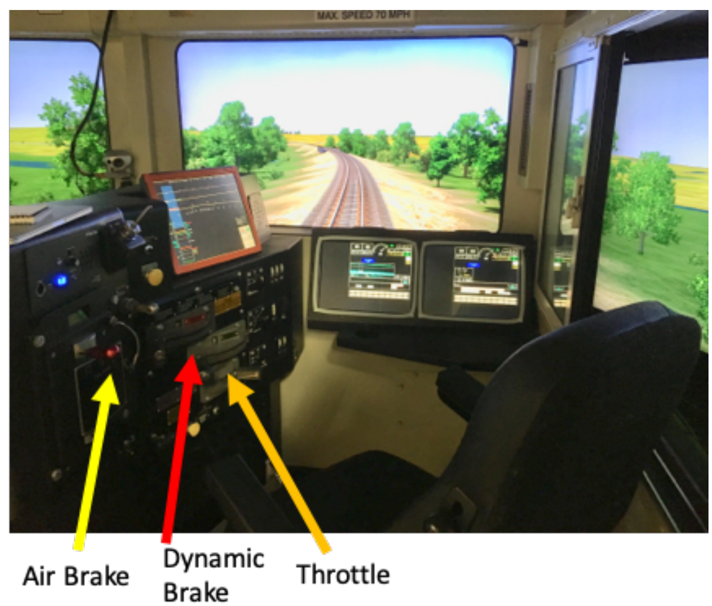
Appendix D

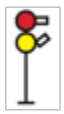
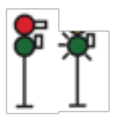
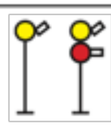
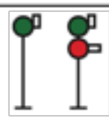
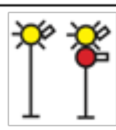
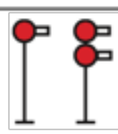
Reference Sheet

The following two pages are the reference sheets for the experiment conducted in the CTIL. The signal chart was provided so that there would be no ambiguity in interpreting the signal meanings, which vary across railroad companies.

- Temporary speed restrictions
  - Will be marked with a blank orange placard two miles before speed restriction is in effect
  - No marking at the actual start of the speed restriction zone
  - Green marker for end of speed restriction zone
- Permanent speed restrictions
  - Will be marked with an orange placard on a diagonal with speed restriction printed on the placard (35 mph, 45 mph, etc) two miles before speed restriction zone
  - The same marked placard sitting parallel to the ground (not on the diagonal) marks the start of the speed restriction zone
  - Green placard marks end of speed restriction





Signal Aspect	Meaning
	**Diverging Approach**. Proceed on diverging route not exceeding prescribed speed through turnout; approach next signal prepared to stop. Trains exceeding 30 mph immediately reduce to that speed.
	**Diverging Clear.** Proceed on diverging route not exceeding prescribed speed through turnout. Note: The flashing green is unusual, but may be used to indicate a diverging clear in this experiment.
	**Approach.** Proceed prepared to stop at next signal. Trains exceeding 30 mph immediately reduce to that speed.
	**Clear.** Proceed.
	**Approach Medium.** Proceed prepared to pass next signal not exceeding 40 mph and be prepared to enter diverging route at prescribed speed.
	**Stop.**

Appendix E

Debrief

The following four pages are the debrief questions for the experiment conducted in the CTIL.

Novice-at-the-Controls: Expert Debrief, Study #___

1. How would you characterize the trial run generally (e.g., track, events, train characteristics, difficulty, etc.)?

2. On a scale of 1-5, how much did you trust the novice to understand your instructions and correctly implement them (1 is minimal trust, 5 is complete trust)? Briefly explain why.

3. Were there any moments during the trip where you did not feel that you could safely control the train or bring it to a stop if necessary? If so, describe the situation(s).

4. Would this have been an acceptable run if you had physically been at the train controls?

5. Do you have any other comments on your performance for this trial?

Novice-at-the-Controls: Novice Debrief, Study #___

1. How would you characterize the trial run generally (e.g., track, events, train characteristics, difficulty, etc.)?

2. On a scale of 1 to 5, how difficult was it for you to a) understand and b) implement the expert's instructions (1 is least difficult, 5 is most difficult)? Briefly explain.

3. Were there any points throughout the trip that you felt the expert's intentions for you were very unclear? If so, describe the situation(s).

4. Do you have any other comments on this trial?

Engineer-at-the-Controls: Expert Debrief, Study #___

1. How would you characterize the trial run generally (e.g., track, events, train characteristics, difficulty, etc.)?

2. On a scale from 1-5, rate how much you trusted the novice to give you accurate updates on the environment (1 is minimal trust, 5 is complete trust). Briefly explain.

3. What was the most difficult aspect of attempting to control the train without being able to see the outside environment?

4. Were there any points during the trial that you were concerned that the train was not being operated in a safe manner? If so, explain the situation(s).

5. Were there any points during the trial that you felt you had inadequate information to operate the train and were unable to obtain the missing information in time? If so, explain the situation(s).

6. Do you have any other comments on this trial?

Engineer-at-the-Controls: Novice Debrief, Study #___

1. How would you characterize the trial run generally (e.g., track, events, train characteristics, difficulty, etc.)?

2. Do you think you gave accurate and timely updates on the environment to the expert? Briefly explain why/why not.

3. Were there situations where you did not understand what information the expert needed or wanted? If so, describe the situation(s).

4. At any point, did you feel like the train was out of control or not being controlled safely given the environment? If so, describe the situation(s) where this occurred.

5. Do you have any other comments on this trial?

Appendix F

Coding Handbook

The following six pages are the coding guide developed for coding the interactions from the experiment.

<u>Level 1:</u>

1. **Request (paraphrased):** Any statement that indicates what the other person should do at a certain moment. Can be a direct command or clarification of the tactic/strategy that is to be followed. It may come off as a suggestion, invitation, favor, request, etc., but it is expected that the other person follows through with it; there is no alternative option to the request.

   a. *<u>Examples</u>*:
      i. "You should be driving at 40 right now."
      ii. "Use more airbrake."

➔ <u>Things to consider</u>: Request and Propose can get mixed up, but the way to tell them apart: Requests have no other alternative, they must be followed; Propose is a suggestion that doesn't necessarily have to be followed. Also, requests are typically given from expert to novice, while propositions are usually given from novice to expert.

2. **Inform:** Any verbalization that communicates information, while not being an inquiry. The utterance could be a response to an inquiry or request, or a notification or observation of cues or aspects of the situation. Any information could also contain a request as part of the utterance. If the information is specifically with regards to a warning, it is explicitly labeled as such.

   a. *<u>Examples</u>*:
      i. "The signal is green."
      ii. "There is a speed restriction in two miles."

➔ <u>Things to consider</u>: Inform can include: warnings, responses to questions, giving context, reminders, etc.

3. **Inquire:** Any verbalization that is a question or inquiry pertaining to the scenario, specific tactics, or information from within or outside the cab.

   a. Query-if (conditional questions)
      i. "If I see x should I do y?"

   b. Query-wh (who, what, why, where, when, how)
      i. "What does the signal say?"
      ii. "When is the speed restriction?"
      iii. "Where are we?"
      iv. "How do I drive this train?"

➔ <u>Things to consider</u>: generally, query-if should be answered with a yes/no, and query-wh is anything else

4. **Acknowledge:** Any verbalization that indicates that the person acknowledges the previous statement, events, or request.
    a. *<u>Examples</u>*:
        i. "Yes."
        ii. "Yeah."
        iii. "Okay."
        iv. "Got it."
        v. (any instance of repeating information as acknowledgement of listening)
5. **Confirm:** Any verbalization that indicates the person agrees with and/or confirms the previous statement, events, or request. A confirmation indicates commitment to a request, or agreement regarding the state of the system and/or environment.
    a. *<u>Examples</u>*:
        i. "Yes, that is green."
        ii. "Yes, we are going downhill."

➔ <u>Things to consider</u>: will typically go with confirm plan in level 2, but not always!

6. **Reject:** Any verbalization that indicates the person disagrees with and/or finds evidence to reject the previous statement, events, or request.
    a. **Implicit**
    b. **Explicit**
        i. "No, it isn't green."
        ii. "No, the back of the train isn't out of the speed restriction."

Level 2:

1. **Check Precondition:** Gathering, looking up, or perceiving information on preconditions (environmental cues, information from paperwork) to prepare for (or that indicates) an upcoming scenario.

➔ Things to consider: usually co-occurs with Remind Plan if it has to do with something that was known before the trip got started, aka something on the track sheet thing (i.e. temporary speed restriction); occurs by itself when someone is merely pointing out a mile post, or pointing out a permanent speed restriction, etc.

➔ Typical Level 1 co-occurrences: Inform

➔ Typical Level 2 co-occurrences: Remind Plan

2. **Remind Plan**: Any reminders of pre-trip planned events, usually referring to temporary speed restrictions or anything already marked up on the paperwork. This is analogous to the conductor calling out reminders of upcoming scenarios. In some instances, this could also be a warning of what might be upcoming or should be expected.

➔ Things to consider: will occur by itself if one person is reminding the other person about a temporary speed restriction (or other occurrence) based on solely looking at the track sheet, and not from a visual cue (mile post, sign, etc.); otherwise, occurs w/ check precondition

➔ Typical Level 1 co-occurrences: Request, Inform

➔ Typical Level 2 co-occurrences: Check Precondition

3. **Confirm Plan**: Any utterance that confirms the upcoming scenario in response to a reminder or a call-out of preconditions.

➔ Things to consider:

➔ Confirm usually responds to: "Is that a signal?" or "Is this a 40 zone?" etc.

➔ Typical Level 1 co-occurrences: Confirm

➔ Typical Level 2 co-occurrences: Usually follows Check Precondition, Remind Plan, or Clarify Plan

4. **Evaluate Plan:** Anything that reconfirms what should be done within the specific context of the scenario. This is usually a response provided to a query reconfirming and validating plan execution.

➔ Things to consider: Can sometimes get mixed up with confirm plan, but the difference is that evaluate plan is when (in most cases) the expert gives his/her opinion on something that (in most cases) the novice brings up. Almost as if expert is saying "yes, that's a good idea."

➔ Evaluate usually responds to: "Should I…..?"

➔ Typical Level 1 co-occurrences: Confirm

- ➔ Typical Level 2 co-occurrences: Explain Plan; also usually follows Clarify, Critique, or Refine Plan

5. **Critique Plan:** When a rationale or execution of a plan is being critiqued or questioned. Can also be a critique of an action or plan that has been executed.

- ➔ Things to consider: Most likely, expert will not be clarifying plan and novice won't be critiquing plan, so this could cause some confusion; there is some inherent bias. So, the same phrase given by novice to expert will probably come off as clarification, while from expert to novice it will come off as critique.
- ➔ Typical Level 1 co-occurrences: Reject, Inquire
- ➔ Typical Level 2 co-occurrences: Clarify Plan

6. **Refine Plan:** When the plan execution is refined or expanded upon with more details or with minor changes based on new information or to current circumstances. The high-level goal remains the same. Most likely to co-occur with Execute Plan, and could occur in response to Critique Plan.

- ➔ Things to consider:
- ➔ Typical Level 1 co-occurrences: Request, Inform
- ➔ Typical Level 2 co-occurrences: Execute Plan, Explain Plan

7. **Clarify Plan:** A question or clarification to ensure the plan execution conforms to expectations and goals. Double-checking and verifying if an action request is being followed correctly. Could also be questions on what should be done given a certain condition.

- ➔ Things to consider: most inquiries will be clarify plan, but not always
- ➔ Typical Level 1 co-occurrences: Inquire, Inform
- ➔ Typical Level 2 co-occurrences: Possibly Critique Plan?

8. **Explain Plan:** The rationale and reason for executing the plan in a certain way. Likely to occur with Explain and Execute Plan because the rationale for why to execute a plan in a certain way is often provided in the same utterance.

- ➔ Things to consider:
- ➔ Typical Level 1 co-occurrences: Inform, sometimes Reject
- ➔ Typical Level 2 co-occurrences: Execute Plan (usually a statement will have an execute plan followed up by explain plan—though you might want to separate into 2 statements), sometimes evaluate or explain

9. **Execute Plan:** What action or set of actions should be employed to execute the plan.

- ➔ Things to consider:
- ➔ Typical Level 1 co-occurrences: Request, Inform, can also in some cases occur w/ Propose
- ➔ Typical Level 2 co-occurrences: Explain plan (same issue as above), Refine Plan, Introduce Plan,

Codes and their abbreviations

Level 1 Code	Symbol	Level 2 Code	Symbol
Acknowledge	ACK	Check Precondition	CKP
Confirm	CON	Clarify Plan	CLF
Discuss	DIS	Confirm Plan	CNF
Inform	INF	Critique Plan	CTQ
Query-If	QIF	Evaluate Plan	EVL
Query-Wh	QWH	Execute Plan	EXC
Reject-Explicit	REX	Refine Plan	REF
		Remind plan	RMD

Subcodes

Controls
air brake
dynamic brake
throttle
notch

Displays
TO display
Accelerometer
Counter/Distance Measuring Device (DMD)

Track
turn out
curve
grades
switch
MP *Y*
track configuration
speed flags (TSR/PSR warning/green flags)
signal (can be by itself or with aspect)
speed restriction (TSR, permanent)
X MPH - refers to speed limit/track speed

Train
train type
slack action
train forces
train breakage
FOT/MOT/EOT (front/middle/end of train)